



## Communities and Crime Unnormalized Data Set

Download: [Data Folder](#). [Data Set Description](#)

**Abstract:** Communities in the US. Data combines socio-economic data from the '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey, and crime data from the 1995 FBI UCR

<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	2215	<b>Area:</b>	Social
<b>Attribute Characteristics:</b>	Real	<b>Number of Attributes:</b>	147	<b>Date Donated</b>	2011-03-02
<b>Associated Tasks:</b>	Regression	<b>Missing Values?</b>	Yes	<b>Number of Web Hits:</b>	58788

### Source:

-- Creator: Michael Redmond (redmond 'at' lasalle.edu); Computer Science; La Salle University; Philadelphia, PA, 19141, USA  
 -- culled from 1990 US Census, 1995 US FBI Uniform Crime Report, 1990 US Law Enforcement Management and Administrative Statistics Survey, available from ICPSR at U of Michigan.  
 -- Donor: Michael Redmond (redmond 'at' lasalle.edu); Computer Science; La Salle University; Philadelphia, PA, 19141, USA

### Data Set Information:

The source datasets needed to be combined via programming. Many variables are included so that algorithms that select or learn weights for attributes could be tested. However, clearly unrelated attributes were not included; attributes were picked if there was any plausible connection to crime (N=125), plus the crime variables which are potential dependent variables. The variables included in the dataset involve the community, such as the percent of the population considered urban, and the median family income, and involving law enforcement, such as per capita number of police officers, and percent of officers assigned to drug units. The crime attributes (N=18) that could be predicted are the 8 crimes considered 'Index Crimes' by the FBI (Murders, Rape, Robbery, ....), per capita (actually per 100,000 population) versions of each, and Per Capita Violent Crimes and Per Capita Nonviolent Crimes).

A limitation was that the LEMAS survey was of the police departments with at least 100 officers, plus a random sample of smaller departments. For our purposes, communities not found in both census and crime datasets were omitted. Many communities are missing LEMAS data.

The per capita crimes variables were calculated using population values included in the 1995 FBI data (which differ from the 1990 Census values).

The per capita violent crimes variable was calculated using population and the sum of crime variables considered violent crimes in the United States: murder, rape, robbery, and assault. There was apparently some controversy in some states concerning the counting of rapes. These resulted in missing values for rape, which resulted in missing values for per capita violent crime. Many of these omitted communities were from the midwestern USA (Minnesota, Illinois, and Michigan have many of these).

The per capita nonviolent crime variable was calculated using the sum of crime variables considered non-violent crimes in the United States: burglaries, larcenies, auto thefts and arsons. (There are many other types of crimes, these only include FBI 'Index Crimes')

Some further pre-processing of the dataset must be done. Choose the desirable dependent variable from among the 18 possible. It would not be interesting or appropriate to predict total crime (e.g. violent crime) while including subtotals (e.g. murders) as independent variables. There are also identifying variables (community name, county code, community code) that are not predictive, and would get in the way of some algorithms. Weka's Unsupervised Attribute Remove Filter can be used to remove unwanted attributes.

The FBI notes that use of this data to evaluate communities is over-simplistic, as many relevant factors are not included. For one example, communities with large numbers of visitors will have higher per capita crime (measured by residents) than communities with fewer visitors, other things being equal.

### Attribute Information:

(125 predictive, 4 non-predictive, 18 potential goal)

-- communityname: Community name - not predictive - for information only (string)  
 -- state: US state (by 2 letter postal abbreviation)(nominal)  
 -- countyCode: numeric code for county - not predictive, and many missing values (numeric)  
 -- communityCode: numeric code for community - not predictive and many missing values (numeric)  
 -- fold: fold number for non-random 10 fold cross validation, potentially useful for debugging, paired tests - not predictive (numeric - integer)

-- population: population for community: (numeric - expected to be integer)  
 -- householdsize: mean people per household (numeric - decimal)  
 -- racePctBlack: percentage of population that is african american (numeric - decimal)  
 -- racePctWhite: percentage of population that is caucasian (numeric - decimal)  
 -- racePctAsian: percentage of population that is of asian heritage (numeric - decimal)  
 -- racePctHispanic: percentage of population that is of hispanic heritage (numeric - decimal)  
 -- agePct12t21: percentage of population that is 12-21 in age (numeric - decimal)  
 -- agePct12t29: percentage of population that is 12-29 in age (numeric - decimal)  
 -- agePct16t24: percentage of population that is 16-24 in age (numeric - decimal)  
 -- agePct65up: percentage of population that is 65 and over in age (numeric - decimal)  
 -- numbUrban: number of people living in areas classified as urban (numeric - expected to be integer)  
 -- pctUrban: percentage of people living in areas classified as urban (numeric - decimal)  
 -- medIncome: median household income (numeric - may be integer)  
 -- pctWWage: percentage of households with wage or salary income in 1989 (numeric - decimal)  
 -- pctWFarmSelf: percentage of households with farm or self employment income in 1989 (numeric - decimal)  
 -- pctWInvInc: percentage of households with investment / rent income in 1989 (numeric - decimal)  
 -- pctWSocSec: percentage of households with social security income in 1989 (numeric - decimal)  
 -- pctWPubAsst: percentage of households with public assistance income in 1989 (numeric - decimal)  
 -- pctWRetire: percentage of households with retirement income in 1989 (numeric - decimal)  
 -- medFamInc: median family income (differs from household income for non-family households) (numeric - may be integer)  
 -- perCapInc: per capita income (numeric - decimal)  
 -- whitePerCap: per capita income for caucasians (numeric - decimal)  
 -- blackPerCap: per capita income for african americans (numeric - decimal)

```

-- indianPerCap: per capita income for native americans (numeric - decimal)
-- AsianPerCap: per capita income for people with asian heritage (numeric - decimal)
-- OtherPerCap: per capita income for people with 'other' heritage (numeric - decimal)
-- HispPerCap: per capita income for people with hispanic heritage (numeric - decimal)
-- NumUnderPov: number of people under the poverty level (numeric - expected to be integer)
-- PctPopUnderPov: percentage of people under the poverty level (numeric - decimal)
-- PctLess9thGrade: percentage of people 25 and over with less than a 9th grade education (numeric - decimal)
-- PctNotHSGrad: percentage of people 25 and over that are not high school graduates (numeric - decimal)
-- PctBSorMore: percentage of people 25 and over with a bachelors degree or higher education (numeric - decimal)
-- PctUnemployed: percentage of people 16 and over, in the labor force, and unemployed (numeric - decimal)
-- PctEmploy: percentage of people 16 and over who are employed (numeric - decimal)
-- PctEmplManu: percentage of people 16 and over who are employed in manufacturing (numeric - decimal)
-- PctEmplProfServ: percentage of people 16 and over who are employed in professional services (numeric - decimal)
-- PctOccupManu: percentage of people 16 and over who are employed in manufacturing (numeric - decimal) ##### No longer sure of difference from PctEmplManu - may
include unemployed manufacturing workers #####
-- PctOccupMgmtProf: percentage of people 16 and over who are employed in management or professional occupations (numeric - decimal)
-- MalePctDivorce: percentage of males who are divorced (numeric - decimal)
-- MalePctNevMarr: percentage of males who have never married (numeric - decimal)
-- FemalePctDiv: percentage of females who are divorced (numeric - decimal)
-- TotalPctDiv: percentage of population who are divorced (numeric - decimal)
-- PersPerFam: mean number of people per family (numeric - decimal)
-- PctFam2Par: percentage of families (with kids) that are headed by two parents (numeric - decimal)
-- PctKids2Par: percentage of kids in family housing with two parents (numeric - decimal)
-- PctYoungKids2Par: percent of kids 4 and under in two parent households (numeric - decimal)
-- PctTeen2Par: percent of kids age 12-17 in two parent households (numeric - decimal)
-- PctWorkMomYoungKids: percentage of moms of kids 6 and under in labor force (numeric - decimal)
-- PctWorkMom: percentage of moms of kids under 18 in labor force (numeric - decimal)
-- NumKidsBornNeverMar: number of kids born to never married (numeric - expected to be integer)
-- PctKidsBornNeverMar: percentage of kids born to never married (numeric - decimal)
-- NumImmig: total number of people known to be foreign born (numeric - expected to be integer)
-- PctImmigRecent: percentage of _immigrants_ who immigrated within last 3 years (numeric - decimal)
-- PctImmigRec5: percentage of _immigrants_ who immigrated within last 5 years (numeric - decimal)
-- PctImmigRec8: percentage of _immigrants_ who immigrated within last 8 years (numeric - decimal)
-- PctImmigRec10: percentage of _immigrants_ who immigrated within last 10 years (numeric - decimal)
-- PctRecentImmig: percent of _population_ who have immigrated within the last 3 years (numeric - decimal)
-- PctRecImmig5: percent of _population_ who have immigrated within the last 5 years (numeric - decimal)
-- PctRecImmig8: percent of _population_ who have immigrated within the last 8 years (numeric - decimal)
-- PctRecImmig10: percent of _population_ who have immigrated within the last 10 years (numeric - decimal)
-- PctSpeakEnglOnly: percent of people who speak only English (numeric - decimal)
-- PctNotSpeakEnglWell: percent of people who do not speak English well (numeric - decimal)
-- PctLargHouseFam: percent of family households that are large (6 or more) (numeric - decimal)
-- PctLargHouseOccup: percent of all occupied households that are large (6 or more people) (numeric - decimal)
-- PersPerOccupHous: mean persons per household (numeric - decimal)
-- PersPerOwnOccHous: mean persons per owner occupied household (numeric - decimal)
-- PersPerRentOccHous: mean persons per rental household (numeric - decimal)
-- PctPersOwnOccup: percent of people in owner occupied households (numeric - decimal)
-- PctPersDenseHous: percent of persons in dense housing (more than 1 person per room) (numeric - decimal)
-- PctHousLess3BR: percent of housing units with less than 3 bedrooms (numeric - decimal)
-- MedNumBR: median number of bedrooms (numeric - decimal)
-- HousVacant: number of vacant households (numeric - expected to be integer)
-- PctHousOccup: percent of housing occupied (numeric - decimal)
-- PctHousOwnOcc: percent of households owner occupied (numeric - decimal)
-- PctVacantBoarded: percent of vacant housing that is boarded up (numeric - decimal)
-- PctVacMore6Mos: percent of vacant housing that has been vacant more than 6 months (numeric - decimal)
-- MedYrHousBuilt: median year housing units built (numeric - may be integer)
-- PctHousNoPhone: percent of occupied housing units without phone (in 1990, this was rare!) (numeric - decimal)
-- PctWOFullPlumb: percent of housing without complete plumbing facilities (numeric - decimal)
-- OwnOccLowQuart: owner occupied housing - lower quartile value (numeric - decimal)
-- OwnOccMedVal: owner occupied housing - median value (numeric - decimal)
-- OwnOccHiQuart: owner occupied housing - upper quartile value (numeric - decimal)
-- OwnOccQrange: owner occupied housing - difference between upper quartile and lower quartile values (numeric - decimal)
-- RentLowQ: rental housing - lower quartile rent (numeric - decimal)
-- RentMedian: rental housing - median rent (Census variable H32B from file STF1A) (numeric - decimal)
-- RentHighQ: rental housing - upper quartile rent (numeric - decimal)
-- RentQrange: rental housing - difference between upper quartile and lower quartile rent (numeric - decimal)
-- MedRent: median gross rent (Census variable H43A from file STF3A - includes utilities) (numeric - decimal)
-- MedRentPctHousInc: median gross rent as a percentage of household income (numeric - decimal)
-- MedOwnCostPctInc: median owners cost as a percentage of household income - for owners with a mortgage (numeric - decimal)
-- MedOwnCostPctIncNoMtg: median owners cost as a percentage of household income - for owners without a mortgage (numeric - decimal)
-- NumInShelters: number of people in homeless shelters (numeric - expected to be integer)
-- NumStreet: number of homeless people counted in the street (numeric - expected to be integer)
-- PctForeignBorn: percent of people foreign born (numeric - decimal)
-- PctBornSameState: percent of people born in the same state as currently living (numeric - decimal)
-- PctSameHouse85: percent of people living in the same house as in 1985 (5 years before) (numeric - decimal)
-- PctSameCity85: percent of people living in the same city as in 1985 (5 years before) (numeric - decimal)
-- PctSameState85: percent of people living in the same state as in 1985 (5 years before) (numeric - decimal)
-- LemasSwornFT: number of sworn full time police officers (numeric - expected to be integer)
-- LemasSwFTPerPop: sworn full time police officers per 100K population (numeric - decimal)
-- LemasSwFTFieldOps: number of sworn full time police officers in field operations (on the street as opposed to administrative etc) (numeric - expected to be integer)
-- LemasSwFTFieldPerPop: sworn full time police officers in field operations (on the street as opposed to administrative etc) per 100K population (numeric - decimal)
-- LemasTotalReq: total requests for police (numeric - expected to be integer)
-- LemasTotReqPerPop: total requests for police per 100K popuation (numeric - decimal)
-- PolicReqPerOffic: total requests for police per police officer (numeric - decimal)
-- PolicPerPop: police officers per 100K population (numeric - decimal)
-- RacialMatchCommPol: a measure of the racial match between the community and the police force. High values indicate proportions in community and police force are similar
(numeric - decimal)
-- PctPolicWhite: percent of police that are caucasian (numeric - decimal)
-- PctPolicBlack: percent of police that are african american (numeric - decimal)
-- PctPolicHisp: percent of police that are hispanic (numeric - decimal)
-- PctPolicAsian: percent of police that are asian (numeric - decimal)
-- PctPolicMinor: percent of police that are minority of any kind (numeric - decimal)
-- OfficAssignDrugUnits: number of officers assigned to special drug units (numeric - expected to be integer)
-- NumKindsDrugsSeiz: number of different kinds of drugs seized (numeric - expected to be integer)
-- PolicAveOTWorked: police average overtime worked (numeric - decimal)
-- LandArea: land area in square miles (numeric - decimal)
-- PopDens: population density in persons per square mile (numeric - decimal)
-- PctUsePubTrans: percent of people using public transit for commuting (numeric - decimal)
-- PolicCars: number of police cars (numeric - expected to be integer)
-- PolicOperBudg: police operating budget (numeric - may be integer)
-- LemasPctPolicOnPatr: percent of sworn full time police officers on patrol (numeric - decimal)
-- LemasGangUnitDeploy: gang unit deployed (numeric - integer - but really nominal - 0 means NO, 10 means YES, 5 means Part Time)
-- LemasPctOfficDrugUn: percent of officers assigned to drug units (numeric - decimal)

```

-- PolicBudgPerPop: police operating budget per population (numeric - decimal)

-- murders: number of murders in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- murdPerPop: number of murders per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- rapes: number of rapes in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- rapesPerPop: number of rapes per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- robberies: number of robberies in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- robbbPerPop: number of robberies per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- assaults: number of assaults in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- assaultPerPop: number of assaults per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- burglaries: number of burglaries in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- burglPerPop: number of burglaries per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- larcenies: number of larcenies in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- larcPerPop: number of larcenies per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- autoTheft: number of auto thefts in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- autoTheftPerPop: number of auto thefts per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- arsons: number of arsons in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)

-- arsonsPerPop: number of arsons per 100K population (numeric - decimal) potential GOAL attribute (to be predicted)

-- ViolentCrimesPerPop: total number of violent crimes per 100K popuation (numeric - decimal) GOAL attribute (to be predicted)

-- nonViolPerPop: total number of non-violent crimes per 100K popuation (numeric - decimal) potential GOAL attribute (to be predicted)

Summary Statistics:  
variable,Minimum,Maximum,Mean,Standard Deviation,Correlation w/ ViolPerPop,Median,Mode,Missing  
pop,10005,7322564,53117.9842,204620.2529,0.212353809,22792,12361,0  
perHoush,1.6,5.28,2.707327314,0.334119654,-0.020110177,2.66,2.6,0  
pctBlack,0.96,67.9,33510158,14.2471563,0.628367756,2.87,0.24,0  
pctWhite,2.68,99.63,83.97981941,16.41908022,-0.676848823,90.35,98.04,0  
pctAsian,0.03,57.46,2.67020316,4.473843444,0.031949477,1.23,0.41,0  
pctHispanic,0.12,95.29,7.950176072,14.58983204,0.253596415,2.18,0.78,0  
pct12-21,4.58,54.4,14.44583747,4.518622844,0.023535103,13.62,13.62,0  
pct12-29,9.38,70.51,27.64483973,6.181517075,0.105908765,26.78,26.78,0  
pct16-24,4.64,63.62,13.97514221,5.970746729,0.045251915,12.54,11.61,0  
pct65up,1.66,52.77,11.83639278,4.777565216,0.055637064,11.73,11.06,0  
persUrban,0.7322564,47734.72009,205606.6933,0.21384036,18041,0,0  
pctUrban,0.100,70.46530926,44.08027532,0.086294263,100,100,0  
medIncome,8866,123625,33984.69616,13424.68001,-0.397407452,31441,27095,0  
pctWwage,31.68,96.76,78.31275847,7.950672435,-0.289829747,78.61,85.12,0  
pctWfarm,0.6,53.0,881841986,0.689005635,-0.148357317,0.69,0.54,0  
pctWdiv,5.81,89.04,43.75093454,12.78792465,-0.557786505,42.88,41.65,0  
pctWsocsec,4.81,76.39,26.40941761,8.295604161,0.107881101,26.59,21.51,0  
pctPubAsst,0.18,44.82,6.801444695,4.700334848,0.563107116,5.61,2.27,0  
pctRetire,3.46,45.51,15.96900226,4.622553011,-0.098964807,15.65,13.14,0  
medFamIncome,10447,139008,39857.05508,14251.20603,-0.411864109,36678,30546,0  
perCapInc,5237,63302,15603.5246,6281.558523,-0.315255977,14101,11252,0  
whitePerCap,5472,68850,16567.69842,6346.840251,-0.185898177,15073,12735,0  
blackPerCap,0.212120,11541.74944,9232.102062,-0.209243037,9777,0,0  
NAperCap,0.480000,12229.19142,14853.83618,-0.060250637,9895,0,0  
asianPerCap,0.106165,14227.98962,9881.266395,-0.128086386,12250,0,0  
otherPerCap,0.137000,9442.765131,7926.466713,-0.103288444,8186,0,1  
hispPerCap,0.54648,11018.99819,5884.063446,-0.219349786,9721,0,0  
persPoverty,78,1384994,7590.853273,39361.46001,0.240253072,2142,470,0  
pctPoverty,0.64,58.11,62053725,8.600352277,0.505349223,9.33,3.26,0  
pctLowEdu,0.2,49.89,9.186645598,6.66670296,0.371421768,7.74,5.78,0  
pctNotHSgrad,1.46,73.66,22.30511964,10.98951654,0.46759552,21.38,11.27,0  
pctCollGrad,1.63,79.18,23.05687585,12.68721342,-0.299897539,19.65,14.2,0  
pctUnemploy,1.32,31.23,6.045241535,2.89561824,0.483440771,5.45,4.36,0  
pctEmploy,24.82,84.67,62.02161174,8.31204498,-0.31764382,62.44,62.6,0  
pctEmployMfg,2.05,50.03,18.22890745,8.099280979,-0.04712343,17.3,25.38,0  
pctEmployProfServ,8.69,62.67,24.53229797,6.659470242,-0.064554813,23.39,21.52,0  
pctOccupManu,1.37,44.27,13.81916479,6.430263859,0.283898507,13.15,16.52,0  
pctOccupMgmt,6.48,64.97,28.2092009,9.326123364,-0.324430512,26.24,28.31,0  
pctMaleDivorc,2.13,20.08,9.12758465,2.802747328,0.510455756,9.15,10.82,0  
pctMaleNevMar,12.06,76.6,30.68351693,8.127990901,0.271229742,29.26,78,0  
pctFemDivorc,3.35,23.92,12.32530023,3.262612867,0.537302801,12.52,14.36,0  
pctAllDivorc,2.83,22.23,10.81251467,3.000883481,0.536548564,10.9,11.77,0  
persPerFam,2.29,4.64,3.129697517,0.240742507,0.149955382,3.1,3.13,0  
pct2Par,22.97,93.6,74.05912867,10.52595184,-0.698640874,75.03,72.16,0  
pctKids2Par,18.3,92.58,71.22725508,12.04504833,-0.728058986,72.53,63.25,0  
pctKids-4w2Par,8.7,100.81,86542212,12.26373565,-0.658436367,83.99,100,0  
pct12-17w2Par,20.2,97.34,75.52178781,10.36526212,-0.655627787,76.92,77.49,0  
pctWorkMom-6,24.42,87.97,60.54264108,8.008936744,-0.02162507,60.71,63.48,0  
pctWorkMom-18,41.95,89.37,68.85479458,6.679959744,-0.146173218,69.23,65.64,0  
kidsBornNevrMarr,0.527557,2141.418962,14692.58284,0.240337327,352,139,0  
pctKidsBornNevrMarr,0.27,35.3,115498871,3.127681493,0.738089083,2.04,0.97,0  
numForeignBorn,20,2082931,6277.273589,55419.65383,0.144585649,1024,147,0  
pctFgnlmmig-3,0.64,29.13,525693,9.780097693,0.156688654,12.26,0,0  
pctFgnlmmig-5,0.76,16.20,42128668,12.41035515,0.201445794,19.08,0,0  
pctFgnlmmig-8,0.80,81.27,54418059,14.36881288,0.236607959,26.72,0,0  
pctFgnlmmig-10,0.88,34.73392777,16.32732237,0.281536984,34.79,0,0  
pctlmmig-3,0.13,71.1,0.99124153,1.595766262,0.215616823,0.5,0,0  
pctlmmig-5,0.19,93.1,697462754,2.46105956,0.232900674,0.75,0,0  
pctlmmig-8,0.25,34.2,307503386,3.286647928,0.23889175,1.04,0,0  
pctlmmig-10,0.32,63.2,943760722,4.246468259,0.250865298,1.31,0,0  
pctSpeakOnlyEng,6.15,98.98,87.07499323,14.07608745,-0.219031839,92.18,93.57,0  
pctNotSpeakEng,0.38,33.2,405792325,4.210367661,0.27243614,0.92,0.44,0  
pctLargHousFam,0.96,34.87,5.38661851,3.794309411,0.341601121,4.28,3.71,0  
pctLargHous,0.44,30.87,3.91578781,3.175770454,0.257215583,3.05,2.98,0  
persPerOccupHous,1.58,4.52,2.615841986,0.315646341,-0.017700909,2.57,2.44,0  
persPerOwnOccup,1.61,4.48,2.74048307,0.297420849,-0.099169625,2.71,2.65,0  
persPerRenterOccup,1.55,4.73,2.367137698,0.391805778,0.240564095,2.29,2.17,0  
pctPersOwnOccup,13.93,97.24,66.36945372,14.18258845,-0.507329566,65.91,63.79,0  
pctPopDenseHous,0.05,59.49,4.132437923,5.599131346,0.395855647,2.34,1.31,0  
pctSmallHousUnits,3.06,95.34,45.40534086,13.77834736,0.454469731,46.39,53.15,0  
medNumBedrm,1.4,2.640632054,0.512686001,-0.347149961,3,3,0  
houseVacant,36,172768,1748.368849,6503.866478,0.289690649,558,246,0  
pctHousOccup,37.47,99.92,93397291,5.04073584,-0.256836226,94.21,95.38,0  
pctHousOwnerOccup,16.86,96.49,63.36829797,13.9700567,-0.45535904,62.83,56.17,0  
pctVacantBoarded,0.39,89.2,778523702,3.592396207,0.479910148,1.66,0,0  
pctVacant6up,3.12,82.13,34.77388713,13.91146771,0.030768767,34.1,37.5,0  
medYrHousBUILT,1939,1987,1962.623476,11.16655501,-0.111201128,1964,1939,0  
pctHousWOphone,0.23,88.4,289823928,4.088174777,0.473717673,2.85,0,0  
pctHousWOplumb,0.5,33.0,425273138,0.426188232,0.311226836,0.32,0,0

ownHousLowQ,14999,500001,88695.80226,66670.78153,-0.194912005,65500,34000,0  
ownHousMed,19500,500001,113097.5233,81906.36228,-0.17822542,82800,500001,0  
ownHousUpperQ,28200,500001,145318.2578,99030.91382,-0.166029142,106700,500001,0  
ownHousQrange,0,331000,56622.45553,39106.49804,-0.086578248,43400,28100,0  
rentLowQ,99,1001,329.9665914,144.1384609,-0.245708342,307.252,0  
rentMed,120,1001,428.537246,170.7066437,-0.232797807,397.316,0  
rentUpperQ,182,1001,527.2528217,199.29078,-0.223893683,486,1001,0  
rentQrange,0,803,197.2862302,85.20568803,-0.110281171,171,139,0  
medGrossRent,192,1001,501.4663657,169.2717347,-0.231754244,467,1001,0  
medRentpctHousInc,14.9,35.1,26.29810384,2.979297132,0.315536838,26.1,24.7,0  
medOwnCostpct,14,32.7,20.99015801,2.987621666,0.061058949,21,22.6,0  
medOwnCostPctWO,10,1.23,4.13,0.01020316,1.419678674,0.063296009,12.8,11.8,0  
persEmergShelt,0,23383,66.95349887,564.253149,0.19494198,0,0,0  
persHomeless,0,10447,17.8234763,245.4525529,0.136446046,0,0,0  
pctForeignBorn,0,18,60.4,7.340302483,8.418475989,0.193229615,4.31,2.97,0  
pctBornStateResid,6,75,93.14,61.5396298,16.75006116,-0.070943862,64.49,74.45,0  
pctSameHouse-5,11.83,78.56,51.53859594,10.51792598,-0.14008735,52.17,54.85,0  
pctSameCounty-5,27.95,96.59,77.41107901,10.87818569,0.082634597,79.49,81.47,0  
pctSameState-5,32.83,99.9,88.11186456,7.287836465,-0.006663834,90.03,92.69,0  
numPolice,65,25655,499.1982507,1681.472251,0.194219827,173,100,1872  
policePerPop,29.4,3437.23,246.4909621,273.7991617,0.073197363,196.01,#N/A,1872  
policeField,14,22496,432.5597668,1493.708385,0.186477279,152,94,1872  
policeFieldPerPop,19.21,3290.62,210.8447813,235.4788145,0.065928304,170.27,183.22,1872  
policeCalls,2100,8328470,252404.9883,689449.7817,0.230023206,90000,50000,1872  
policeCallPerPop,2704.8,1926281.5,120651.7189,148211.3422,0.14854896,91034.6,#N/A,1872  
policeCallPerOffic,20.8,2162.5,523.658309,307.8390067,0.145798061,443.2,422.6,1872  
policePerPop2,29.4,3437.2,246.493586,273.7984086,0.073203166,196.171,5.1872  
racialMatch,42.15,100,85.4996793,10.94131216,-0.469374164,87.93,100,1872  
pctPoliceWhite,1.6,100,82.5158309,15.33261214,-0.392584252,86.18,100,1872  
pctPoliceBlack,0,67.31,9.263294461,11.02142376,0.513568853,5,0,1872  
pctPoliceHisp,0,98.4,5.459766764,10.60453332,0.056080974,2.04,0,1872  
pctPoliceAsian,0,18.57,0.681282799,1.706344058,0.072778715,0,0,1872  
pctPoliceMinority,0,98.4,15.2422449,14.82675626,0.416418453,11.37,0,1872  
officDrugUnits,0,1773,26.28862974,100.8219209,0.166934688,12.6,1872  
numDiffDrugsSeiz,1,15,8.816326531,2.836390802,0.115940631,9.9,1872  
policeAveOT,0,634.7,119.1142857,92.49518559,0.010162351,98.7,0,1872  
landArea,0,9,3569.8,27.41995485,109.8226001,0.075697852,13.7,4.9,0  
popDensity,10,44229.9,2783.835034,2828.993341,0.256966815,2027.3,3217.7,0  
pctUsePubTrans,0,54.33,3.041124153,4.91291686,0.190478991,1.22,0,0  
policeCarsAvail,20,3187,185.4781341,318.5428335,0.313164478,86,55,1872  
policeOperBudget,2380215,1617293056,32176019.34,110456627.5,0.195267289,11164110,8000000,1872  
pctPolicePatrol,10.85,99.94,87.13093294,10.34961235,-0.093114808,89.58,93.07,1872  
gangUnit,0,10,4.285714286,4.064537838,0.109156269,5,0,1872  
pctOfficDrugUnit,0,48.44,0.980162528,2.877127691,0.318474028,0,0,0  
policeBudgetPerPop,15260.4,2422367.153577.8712,203040.8861,0.056100531,114582,#N/A,1872  
murders,0,1946,7.764785553,58.16646847,0.248259306,1,0,0  
murderPerPop,0,91.09,5.859295711,9.156828742,0.671541352,2.17,0,0  
rapes,0,2818,28.04633782,105.6161353,0.336087881,7,0,208  
rapesPerPop,0,401.35,36.25848032,34.23974957,0.581533276,26.92,0,208  
robberies,0,86001,237.9521229,2250.720788,0.208547391,19,1,1  
robberyPerPop,0,2264.13,162.6125971,234.4866243,0.828574083,74.8,0,1  
assaults,0,62778,326.5281562,1987.947941,0.301016534,56,12,13  
assaultPerPop,0,4932.5,378.0046049,438.2385994,0.945565838,226.525,0,13  
burglaries,2,99207,761.2368897,3111.702756,0.316504514,205,79,3  
burglaryPerPop,16.92,11881.02,1033.430203,763.3544416,0.698552626,822.715,728.93,3  
larcenies,10,235132,2137.629295,7600.573464,0.2950221,747,547,3  
larcPerPop,77.86,25910.55,3372.97915,1901.316145,0.509410495,3079.51,4631.1,3  
autoTheft,1,112464,516.6925859,3258.164244,0.244925811,75,16,3  
autoTheftPerPop,6.55,4968.59,473.9656284,504.6660256,0.636484339,302.355,213.62,3  
arsons,0,5119,30.90772128,180.1252481,0.232824758,5,0,91  
arsonsPerPop,0,436.37,32.15368173,39.24090028,0.416718515,21.08,0,91  
violentPerPop,0,4877.06,589.0789218,614.7845182,1.374,0.6,223.06,221  
nonViolPerPop,116.79,27119.76,4908.241804,2739.708901,0.675374243,4425.45,4295.96,97

Statistics for nominal State variable:  
State,Count  
AK,3  
AL,43  
AR,25  
AZ,20  
CA,279  
CO,25  
CT,71  
DC,1  
DE,1  
FL,90  
GA,37  
IA,20  
ID,7  
IL,40  
IN,48  
KS,1  
KY,26  
LA,22  
MA,123  
MD,12  
ME,17  
MI,108  
MN,66  
MO,42  
MS,20  
NC,46  
ND,8  
NH,21  
NJ,211  
NM,10  
NV,5  
NY,46  
OH,111  
OK,36  
OR,31  
PA,101



RI,26  
SC,28  
SD,9  
TN,35  
TX,162  
UT,24  
VA,33  
VT,4  
WA,40  
WI,60  
WV,14  
WY,7

Distribution of the main Goal Variable (Violent Crimes per Population):  
Range Frequency (On boundary goes in the lower bin; e.g. exactly 200 goes in 100-200)

0,1  
1-100,285  
100-200,306  
200-300,265  
300-400,185  
400-500,151  
500-600,131  
600-700,100  
700-800,77  
800-900,72  
900-1000,61  
1000-1100,38  
1100-1200,33  
1200-1300,50  
1300-1400,35  
1400-1500,30  
1500-1600,28  
1600-1700,28  
1700-1800,14  
1800-1900,12  
1900-2000,14  
More,78

.arff header for weka:

@relation crimeunnormalized

@attribute communityname string  
@attribute State  
{AK,AL,AR,AZ,CA,CO,CT,DC,DE,FL,GA,IA,ID,IL,IN,KS,KY,LA,MA,MD,ME,MI,MN,MO,MS,NC,ND,NH,NJ,NM,NV,NY,OH,OK,OR,PA,RI,SC,SD,TN,TX,UT,VA,VT,WA,WI,WV,WY}  
@attribute countyCode numeric  
@attribute communityCode numeric  
@attribute fold numeric  
@attribute pop numeric  
@attribute perHoush numeric  
@attribute pctBlack numeric  
@attribute pctWhite numeric  
@attribute pctAsian numeric  
@attribute pctHisp numeric  
@attribute pct12-21 numeric  
@attribute pct12-29 numeric  
@attribute pct16-24 numeric  
@attribute pct65Up numeric  
@attribute persUrban numeric  
@attribute pctUrban numeric  
@attribute medIncome numeric  
@attribute pctWwage numeric  
@attribute pctWfarm numeric  
@attribute pctWdiv numeric  
@attribute pctWsocsec numeric  
@attribute pctPubAsst numeric  
@attribute pctRetire numeric  
@attribute medFamIncome numeric  
@attribute perCapInc numeric  
@attribute whitePerCap numeric  
@attribute blackPerCap numeric  
@attribute NAPERCap numeric  
@attribute asianPerCap numeric  
@attribute otherPerCap numeric  
@attribute hispPerCap numeric  
@attribute persPoverty numeric  
@attribute pctPoverty numeric  
@attribute pctLowEdu numeric  
@attribute pctNotHSgrad numeric  
@attribute pctCollGrad numeric  
@attribute pctUnemploy numeric  
@attribute pctEmploy numeric  
@attribute pctEmployMfg numeric  
@attribute pctEmployProfServ numeric  
@attribute pctOccupManu numeric  
@attribute pctOccupMgmt numeric  
@attribute pctMaleDivorc numeric  
@attribute pctMaleNevMar numeric  
@attribute pctFemDivorc numeric  
@attribute pctAllDivorc numeric  
@attribute persPerFam numeric  
@attribute pct2Par numeric  
@attribute pctKids2Par numeric  
@attribute pctKids-4w2Par numeric  
@attribute pct12-17w2Par numeric  
@attribute pctWorkMom-6 numeric  
@attribute pctWorkMom-18 numeric  
@attribute kidsBornNevrMarr numeric  
@attribute pctKidsBornNevrMarr numeric  
@attribute numForeignBorn numeric  
@attribute pctFgnImmig-3 numeric

@attribute pctFgnImmig-5 numeric  
 @attribute pctFgnImmig-8 numeric  
 @attribute pctFgnImmig-10 numeric  
 @attribute pctImmig-3 numeric  
 @attribute pctImmig-5 numeric  
 @attribute pctImmig-8 numeric  
 @attribute pctImmig-10 numeric  
 @attribute pctSpeakOnlyEng numeric  
 @attribute pctNotSpeakEng numeric  
 @attribute pctLargHousFam numeric  
 @attribute pctLargHous numeric  
 @attribute persPerOccupHous numeric  
 @attribute persPerOwnOccup numeric  
 @attribute persPerRenterOccup numeric  
 @attribute pctPersOwnOccup numeric  
 @attribute pctPopDenseHous numeric  
 @attribute pctSmallHousUnits numeric  
 @attribute medNumBedrm numeric  
 @attribute houseVacant numeric  
 @attribute pctHousOccup numeric  
 @attribute pctHousOwnerOccup numeric  
 @attribute pctVacantBoarded numeric  
 @attribute pctVacant6up numeric  
 @attribute medYrHousBuilt numeric  
 @attribute pctHousWOphone numeric  
 @attribute pctHousWOplumb numeric  
 @attribute ownHousLowQ numeric  
 @attribute ownHousMed numeric  
 @attribute ownHousUpperQ numeric  
 @attribute ownHousQrange numeric  
 @attribute rentLowQ numeric  
 @attribute rentMed numeric  
 @attribute rentUpperQ numeric  
 @attribute rentQrange numeric  
 @attribute medGrossRent numeric  
 @attribute medRentpctHousInc numeric  
 @attribute medOwnCostpct numeric  
 @attribute medOwnCostPctWO numeric  
 @attribute persEmergShelt numeric  
 @attribute persHomeless numeric  
 @attribute pctForeignBorn numeric  
 @attribute pctBornStateResid numeric  
 @attribute pctSameHouse-5 numeric  
 @attribute pctSameCounty-5 numeric  
 @attribute pctSameState-5 numeric  
 @attribute numPolice numeric  
 @attribute policePerPop numeric  
 @attribute policeField numeric  
 @attribute policeFieldPerPop numeric  
 @attribute policeCalls numeric  
 @attribute policCallPerPop numeric  
 @attribute policCallPerOffic numeric  
 @attribute policePerPop2 numeric  
 @attribute racialMatch numeric  
 @attribute pctPolicWhite numeric  
 @attribute pctPolicBlack numeric  
 @attribute pctPolicHisp numeric  
 @attribute pctPolicAsian numeric  
 @attribute pctPolicMinority numeric  
 @attribute officDrugUnits numeric  
 @attribute numDiffDrugsSeiz numeric  
 @attribute policAveOT numeric  
 @attribute landArea numeric  
 @attribute popDensity numeric  
 @attribute pctUsePubTrans numeric  
 @attribute policCarsAvail numeric  
 @attribute policOperBudget numeric  
 @attribute pctPolicPatrol numeric  
 @attribute gangUnit numeric  
 @attribute pctOfficDrugUnit numeric  
 @attribute policBudgetPerPop numeric  
 @attribute murders numeric  
 @attribute murdPerPop numeric  
 @attribute rapes numeric  
 @attribute rapesPerPop numeric  
 @attribute robberies numeric  
 @attribute robbbPerPop numeric  
 @attribute assaults numeric  
 @attribute assaultPerPop numeric  
 @attribute burglaries numeric  
 @attribute burglPerPop numeric  
 @attribute larcenies numeric  
 @attribute larcPerPop numeric  
 @attribute autoTheft numeric  
 @attribute autoTheftPerPop numeric  
 @attribute arsons numeric  
 @attribute arsonsPerPop numeric  
 @attribute violentPerPop numeric  
 @attribute nonViolPerPop numeric  
 @data

## Relevant Papers:

### Past Usage

1. [Redmond and Highley 2009] Redmond, M., and Highley, T., Empirical Analysis of Case-Editing Approaches for Numeric Prediction. In International Joint Conference on Computer, Information, and Systems Sciences and Engineering (CISSE) subconference International Conference on Systems, Computing Sciences and Software Engineering (SCSS). University of Bridgeport, CT, December 2009.

-- All numeric data was normalized (0-1), ViolentCrimesPerPop was predicted (all other crime attributes were eliminated)

-- Best mean absolute error obtained was .096 (on normalized data)

2. [Buczak and Gifford 2010] Buczak, A. L. and Gifford, C. M., Fuzzy Association Rule Mining for Community Crime Pattern Discovery. In Workshop on Intelligence and Security Informatics at 16th Conference on Knowledge Discovery and Data Mining (ISI-KDD-2010). Washington DC. July 2010.

-- Data was further processed

### Citation Request:

Please cite the UCI Machine Learning Repository and my sources:

U. S. Department of Commerce, Bureau of the Census, Census Of Population And Housing 1990 United States: Summary Tape File 1a & 3a (Computer Files),

U.S. Department Of Commerce, Bureau Of The Census Producer, Washington, DC and Inter-university Consortium for Political and Social Research Ann Arbor, Michigan. (1992)

U.S. Department of Justice, Bureau of Justice Statistics, Law Enforcement Management And Administrative Statistics (Computer File) U.S. Department Of Commerce, Bureau Of The Census Producer, Washington, DC and Inter-university Consortium for Political and Social Research Ann Arbor, Michigan. (1992)

U.S. Department of Justice, Federal Bureau of Investigation, Crime in the United States (Computer File) (1995)

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)