

Missing Data Treatment and Outlier Analysis

Lujing Xi

6/5/2021

```
library(readr)
library(readxl)

crime <- read.csv("crimedata.csv", na = "?")

head(crime)
```

```
##      Ecommunityname state countyCode communityCode fold population
## 1 BerkeleyHeightstownship NJ      39      5320      1      11980
## 2 Marpletownship PA      45      47616      1      23123
## 3 Tigardcity OR      NA      NA      1      29344
## 4 Gloversvillecity NY      35      29443      1      16656
## 5 Bemidjicity MN      7      5068      1      11245
## 6 Springfieldcity MO      NA      NA      1      140494
##      householdsize racePctblack racePctWhite racePctAsian racePctHispanic agePct12t21
## 1      3.10      1.37      91.78      6.50      1.88      12.47
## 2      2.82      0.80      95.57      3.44      0.85      11.01
## 3      2.43      0.74      94.33      3.43      2.35      11.36
## 4      2.40      1.70      97.35      0.50      0.70      12.55
## 5      2.76      0.53      89.16      1.17      0.52      24.46
## 6      2.45      2.51      95.65      0.90      0.95      18.09
##      agePct12t29 agePct16t24 agePct65up numbUrban pctUrban medIncome pctWWage
## 1      21.44      10.93      11.33      11980      100      75122      89.24
## 2      21.30      10.48      17.18      23123      100      47917      78.99
## 3      25.88      11.01      10.28      29344      100      35669      82.00
## 4      25.20      12.19      17.57      0      0      20580      68.15
## 5      40.53      28.69      12.65      0      0      17390      69.33
## 6      32.89      20.04      13.26      140494      100      21577      75.78
##      pctWFarmSelf pctWInvInc pctWSocSec pctWPubAsst pctWRetire medFamInc perCapInc
## 1      1.55      70.20      23.62      1.03      18.39      79584      29711
## 2      1.11      64.11      35.50      2.75      22.85      55323      20148
## 3      1.15      55.73      22.25      2.94      14.56      42112      16946
## 4      0.24      38.95      39.48      11.71      18.33      26501      10810
## 5      0.55      42.82      32.16      11.21      14.43      24018      8483
## 6      1.00      41.15      29.31      7.12      14.09      27705      11878
##      whitePerCap blackPerCap indianPerCap AsianPerCap OtherPerCap HispPerCap
## 1      30233      13600      5725      27101      5115      22838
## 2      20191      18137      0      20074      5250      12222
## 3      17103      16644      21606      15528      5954      8405
## 4      10909      9984      4941      3541      2451      4391
## 5      9009      887      4425      3352      3000      1328
```

## 6	12029	7382	10264	10753	7192	8104
##	NumUnderPov	PctPopUnderPov	PctLess9thGrade	PctNotHSGrad	PctBSorMore	
## 1	227	1.96	5.81	9.90	48.18	
## 2	885	3.98	5.61	13.72	29.89	
## 3	1389	4.75	2.80	9.09	30.13	
## 4	2831	17.23	11.05	33.68	10.81	
## 5	2855	29.99	12.15	23.06	25.28	
## 6	23223	17.78	8.76	23.03	20.66	
##	PctUnemployed	PctEmploy	PctEmplManu	PctEmplProfServ	PctOccupManu	
## 1	2.70	64.55	14.65	28.82	5.49	
## 2	2.43	61.96	12.26	29.28	6.39	
## 3	4.01	69.80	15.95	21.52	8.79	
## 4	9.86	54.74	31.22	27.43	26.76	
## 5	9.08	52.44	6.89	36.54	10.94	
## 6	5.72	59.02	14.31	26.83	14.72	
##	PctOccupMgmtProf	MalePctDivorce	MalePctNevMarr	FemalePctDiv	TotalPctDiv	
## 1	50.73	3.67	26.38	5.22	4.47	
## 2	37.64	4.23	27.99	6.45	5.42	
## 3	32.48	10.10	25.78	14.76	12.55	
## 4	22.71	10.98	28.15	14.47	12.91	
## 5	27.80	7.51	50.66	11.64	9.73	
## 6	23.42	11.40	33.32	14.46	13.04	
##	PersPerFam	PctFam2Par	PctKids2Par	PctYoungKids2Par	PctTeen2Par	
## 1	3.22	91.43	90.17	95.78	95.81	
## 2	3.11	86.91	85.33	96.82	86.46	
## 3	2.95	78.54	78.85	92.37	75.72	
## 4	2.98	64.02	62.36	65.38	67.43	
## 5	2.98	58.59	55.20	66.51	79.17	
## 6	2.89	71.94	69.79	79.76	75.33	
##	PctWorkMomYoungKids	PctWorkMom	NumKidsBornNeverMar	PctKidsBornNeverMar		
## 1	44.56	58.88	31	0.36		
## 2	51.14	62.43	43	0.24		
## 3	66.08	74.19	164	0.88		
## 4	59.59	70.27	561	3.84		
## 5	61.22	68.94	402	4.70		
## 6	62.96	70.52	1511	1.58		
##	NumImmig	PctImmigRecent	PctImmigRec5	PctImmigRec8	PctImmigRec10	
## 1	1277	8.69	13.00	20.99	30.93	
## 2	1920	5.21	8.65	13.33	22.50	
## 3	1468	16.42	23.98	32.08	35.63	
## 4	339	13.86	13.86	15.34	15.34	
## 5	196	46.94	56.12	67.86	69.90	
## 6	2091	21.33	30.56	38.02	45.48	
##	PctRecentImmig	PctRecImmig5	PctRecImmig8	PctRecImmig10	PctSpeakEnglOnly	
## 1	0.93	1.39	2.24	3.30	85.68	
## 2	0.43	0.72	1.11	1.87	87.79	
## 3	0.82	1.20	1.61	1.78	93.11	
## 4	0.28	0.28	0.31	0.31	94.98	
## 5	0.82	0.98	1.18	1.22	94.64	
## 6	0.32	0.45	0.57	0.68	96.87	
##	PctNotSpeakEnglWell	PctLargHouseFam	PctLargHouseOccup	PersPerOccupHous		
## 1	1.37	4.81	4.17	2.99		
## 2	1.81	4.25	3.34	2.70		
## 3	1.14	2.97	2.05	2.42		

## 4	0.56	3.93	2.56	2.37		
## 5	0.39	5.23	3.11	2.35		
## 6	0.60	3.08	1.92	2.28		
##	PersPerOwnOccHous	PersPerRentOccHous	PctPersOwnOccup	PctPersDenseHous		
## 1	3.00	2.84	91.46	0.39		
## 2	2.83	1.96	89.03	1.01		
## 3	2.69	2.06	64.18	2.03		
## 4	2.51	2.20	58.18	1.21		
## 5	2.55	2.12	58.13	2.94		
## 6	2.37	2.16	57.81	2.11		
##	PctHousLess3BR	MedNumBR	HousVacant	PctHousOccup	PctHousOwnOcc	
## 1	11.06	3	64	98.37	91.01	
## 2	23.60	3	240	97.15	84.88	
## 3	47.46	3	544	95.68	57.79	
## 4	45.66	3	669	91.19	54.89	
## 5	55.64	2	333	92.45	53.57	
## 6	53.19	2	5119	91.81	55.50	
##	PctVacantBoarded	PctVacMore6Mos	MedYrHousBuilt	PctHousNoPhone	PctWOFullPlumb	
## 1	3.12	37.50	1959	0.00	0.28	
## 2	0.00	18.33	1958	0.31	0.14	
## 3	0.92	7.54	1976	1.55	0.12	
## 4	2.54	57.85	1939	7.00	0.87	
## 5	3.90	42.64	1958	7.45	0.82	
## 6	2.09	26.22	1966	6.13	0.31	
##	OwnOccLowQuart	OwnOccMedVal	OwnOccHiQuart	OwnOccQrange	RentLowQ	RentMedian
## 1	215900	262600	326900	111000	685	1001
## 2	136300	164200	199900	63600	467	560
## 3	74700	90400	112000	37300	370	428
## 4	36400	49600	66500	30100	195	250
## 5	30600	43200	59500	28900	202	283
## 6	37700	53900	73100	35400	215	280
##	RentHighQ	RentQrange	MedRent	MedRentPctHousInc	MedOwnCostPctInc	
## 1	1001	316	1001	23.8	21.1	
## 2	672	205	627	27.6	20.7	
## 3	520	150	484	24.1	21.7	
## 4	309	114	333	28.7	20.6	
## 5	362	160	332	32.2	23.2	
## 6	349	134	340	26.4	17.3	
##	MedOwnCostPctIncNoMtg	NumInShelters	NumStreet	PctForeignBorn	PctBornSameState	
## 1	14.0	11	0	10.66	53.72	
## 2	12.5	0	0	8.30	77.17	
## 3	11.6	16	0	5.00	44.77	
## 4	14.5	0	0	2.04	88.71	
## 5	12.9	2	0	1.74	73.75	
## 6	11.7	327	4	1.49	64.35	
##	PctSameHouse85	PctSameCity85	PctSameState85	LemasSwornFT	LemasSwFTPerPop	
## 1	65.29	78.09	89.14	NA	NA	
## 2	71.27	90.22	96.12	NA	NA	
## 3	36.60	61.26	82.85	NA	NA	
## 4	56.70	90.17	96.24	NA	NA	
## 5	42.22	60.34	89.02	NA	NA	
## 6	42.29	70.61	85.66	NA	NA	
##	LemasSwFTFieldOps	LemasSwFTFieldPerPop	LemasTotalReq	LemasTotReqPerPop		
## 1	NA	NA	NA	NA		

## 2	NA	NA	NA	NA			
## 3	NA	NA	NA	NA			
## 4	NA	NA	NA	NA			
## 5	NA	NA	NA	NA			
## 6	NA	NA	NA	NA			
##	PolicReqPerOffic	PolicPerPop	RacialMatchCommPol	PctPolicWhite	PctPolicBlack		
## 1	NA	NA	NA	NA	NA		
## 2	NA	NA	NA	NA	NA		
## 3	NA	NA	NA	NA	NA		
## 4	NA	NA	NA	NA	NA		
## 5	NA	NA	NA	NA	NA		
## 6	NA	NA	NA	NA	NA		
##	PctPolicHisp	PctPolicAsian	PctPolicMinor	OfficAssgnDrugUnits			
## 1	NA	NA	NA	NA			
## 2	NA	NA	NA	NA			
## 3	NA	NA	NA	NA			
## 4	NA	NA	NA	NA			
## 5	NA	NA	NA	NA			
## 6	NA	NA	NA	NA			
##	NumKindsDrugsSeiz	PolicAveOTWorked	LandArea	PopDens	PctUsePubTrans	PolicCars	
## 1	NA	NA	6.5	1845.9	9.63	NA	
## 2	NA	NA	10.6	2186.7	3.84	NA	
## 3	NA	NA	10.6	2780.9	4.37	NA	
## 4	NA	NA	5.2	3217.7	3.31	NA	
## 5	NA	NA	11.5	974.2	0.38	NA	
## 6	NA	NA	70.4	1995.7	0.97	NA	
##	PolicOperBudg	LemasPctPolicOnPatr	LemasGangUnitDeploy	LemasPctOfficDrugUn			
## 1	NA	NA	NA	0			
## 2	NA	NA	NA	0			
## 3	NA	NA	NA	0			
## 4	NA	NA	NA	0			
## 5	NA	NA	NA	0			
## 6	NA	NA	NA	0			
##	PolicBudgPerPop	murders	murdPerPop	rapes	rapesPerPop	robberies	robberPerPop
## 1	NA	0	0.00	0	0.00	1	8.20
## 2	NA	0	0.00	1	4.25	5	21.26
## 3	NA	3	8.30	6	16.60	56	154.95
## 4	NA	0	0.00	10	57.86	10	57.86
## 5	NA	0	0.00	NA	NA	4	32.04
## 6	NA	7	4.63	77	50.98	136	90.05
##	assaults	assaultPerPop	burglaries	burglPerPop	larcenies	larcPerPop	autoTheft
## 1	4	32.81	14	114.85	138	1132.08	16
## 2	24	102.05	57	242.37	376	1598.78	26
## 3	14	38.74	274	758.14	1797	4972.19	136
## 4	33	190.93	225	1301.78	716	4142.56	47
## 5	14	112.14	91	728.93	1060	8490.87	91
## 6	449	297.29	2094	1386.46	7690	5091.64	454
##	autoTheftPerPop	arsons	arsonsPerPop	ViolentCrimesPerPop	nonViolPerPop		
## 1	131.26	2	16.41	41.02	1394.59		
## 2	110.55	1	4.25	127.56	1955.95		
## 3	376.30	22	60.87	218.59	6167.51		
## 4	271.93	NA	NA	306.64	NA		
## 5	728.93	5	40.05	NA	9988.79		
## 6	300.60	134	88.72	442.95	6867.42		

```
# In the original data set, all missing values
# were represented by '?'. When loading data, I
# added na='?' to recode ? into NA.
```

```
summary(crime)
```

```
##      Ecommunityname      state      countyCode      communityCode
## Length:2215      Length:2215      Min.   : 1.00      Min.   : 70
## Class :character      Class :character      1st Qu.: 11.00      1st Qu.:22887
## Mode  :character      Mode  :character      Median : 27.00      Median :46925
##                                           Mean  : 65.59      Mean  :45209
##                                           3rd Qu.: 80.50      3rd Qu.:65805
##                                           Max.   :840.00      Max.   :94597
##                                           NA's   :1221      NA's   :1224
##      fold      population      householdsize      racepctblack
## Min.   : 1.000      Min.   : 10005      Min.   :1.600      Min.   : 0.000
## 1st Qu.: 3.000      1st Qu.: 14366      1st Qu.:2.500      1st Qu.: 0.860
## Median : 5.000      Median : 22792      Median :2.660      Median : 2.870
## Mean   : 5.494      Mean   : 53118      Mean   :2.707      Mean   : 9.335
## 3rd Qu.: 8.000      3rd Qu.: 43024      3rd Qu.:2.850      3rd Qu.:11.145
## Max.   :10.000      Max.   :7322564      Max.   :5.280      Max.   :96.670
##
##      racePctWhite      racePctAsian      racePctHisp      agePct12t21
## Min.   : 2.68      Min.   : 0.03      Min.   : 0.12      Min.   : 4.58
## 1st Qu.:76.32      1st Qu.: 0.62      1st Qu.: 0.93      1st Qu.:12.25
## Median :90.35      Median : 1.23      Median : 2.18      Median :13.62
## Mean   :83.98      Mean   : 2.67      Mean   : 7.95      Mean   :14.45
## 3rd Qu.:96.22      3rd Qu.: 2.67      3rd Qu.: 7.81      3rd Qu.:15.36
## Max.   :99.63      Max.   :57.46      Max.   :95.29      Max.   :54.40
##
##      agePct12t29      agePct16t24      agePct65up      numbUrban
## Min.   : 9.38      Min.   : 4.64      Min.   : 1.66      Min.   : 0
## 1st Qu.:24.41      1st Qu.:11.32      1st Qu.: 8.75      1st Qu.: 0
## Median :26.78      Median :12.54      Median :11.73      Median : 18041
## Mean   :27.64      Mean   :13.98      Mean   :11.84      Mean   : 47735
## 3rd Qu.:29.20      3rd Qu.:14.35      3rd Qu.:14.41      3rd Qu.: 41918
## Max.   :70.51      Max.   :63.62      Max.   :52.77      Max.   :7322564
##
##      pctUrban      medIncome      pctWWage      pctWFarmSelf
## Min.   : 0.00      Min.   : 8866      Min.   :31.68      Min.   :0.0000
## 1st Qu.: 0.00      1st Qu.: 23817      1st Qu.:73.40      1st Qu.:0.4600
## Median :100.00      Median : 31441      Median :78.61      Median :0.6900
## Mean   : 70.47      Mean   : 33985      Mean   :78.31      Mean   :0.8818
## 3rd Qu.:100.00      3rd Qu.: 41481      3rd Qu.:84.03      3rd Qu.:1.1000
## Max.   :100.00      Max.   :123625      Max.   :96.76      Max.   :6.5300
##
##      pctWInvInc      pctWSocSec      pctWPubAsst      pctWRetire
## Min.   : 5.81      Min.   : 4.81      Min.   : 0.180      Min.   : 3.46
## 1st Qu.:34.68      1st Qu.:20.77      1st Qu.: 3.270      1st Qu.:12.88
## Median :42.88      Median :26.59      Median : 5.610      Median :15.65
## Mean   :43.75      Mean   :26.41      Mean   : 6.801      Mean   :15.97
## 3rd Qu.:52.74      3rd Qu.:31.71      3rd Qu.: 9.105      3rd Qu.:18.75
## Max.   :89.04      Max.   :76.39      Max.   :44.820      Max.   :45.51
```

```

##
##      medFamInc      perCapInc      whitePerCap      blackPerCap
## Min.   : 10447    Min.   : 5237    Min.   : 5472    Min.   :    0
## 1st Qu.: 29538    1st Qu.:11602    1st Qu.:12610    1st Qu.: 6742
## Median : 36678    Median :14101    Median :15073    Median : 9777
## Mean   : 39857    Mean   :15604    Mean   :16568    Mean   :11542
## 3rd Qu.: 46999    3rd Qu.:17795    3rd Qu.:18610    3rd Qu.:14526
## Max.   :139008    Max.   :63302    Max.   :68850    Max.   :212120
##
##      indianPerCap      AsianPerCap      OtherPerCap      HispPerCap
## Min.   :    0    Min.   :    0    Min.   :    0    Min.   :    0
## 1st Qu.: 6345    1st Qu.: 8286    1st Qu.: 5528    1st Qu.: 7274
## Median : 9895    Median :12250    Median : 8186    Median : 9721
## Mean   :12229    Mean   :14228    Mean   : 9443    Mean   :11019
## 3rd Qu.:14758    3rd Qu.:17328    3rd Qu.:11526    3rd Qu.:13418
## Max.   :480000    Max.   :106165    Max.   :137000    Max.   :54648
##
##                                     NA's      :1
##      NumUnderPov      PctPopUnderPov      PctLess9thGrade      PctNotHSGrad
## Min.   :    78.0    Min.   : 0.64    Min.   : 0.200    Min.   : 1.46
## 1st Qu.:   912.5    1st Qu.: 4.51    1st Qu.: 4.640    1st Qu.:13.92
## Median :  2142.0    Median : 9.33    Median : 7.740    Median :21.38
## Mean   :  7590.9    Mean   :11.62    Mean   : 9.187    Mean   :22.31
## 3rd Qu.:  4988.0    3rd Qu.:16.91    3rd Qu.:11.835    3rd Qu.:29.20
## Max.   :1384994.0    Max.   :58.00    Max.   :49.890    Max.   :73.66
##
##      PctBSorMore      PctUnemployed      PctEmploy      PctEmplManu
## Min.   : 1.63    Min.   : 1.320    Min.   :24.82    Min.   : 2.05
## 1st Qu.:14.10    1st Qu.: 4.045    1st Qu.:56.49    1st Qu.:12.21
## Median :19.65    Median : 5.450    Median :62.44    Median :17.30
## Mean   :23.06    Mean   : 6.045    Mean   :62.02    Mean   :18.23
## 3rd Qu.:29.05    3rd Qu.: 7.440    3rd Qu.:67.83    3rd Qu.:23.40
## Max.   :79.18    Max.   :31.230    Max.   :84.67    Max.   :50.03
##
##      PctEmplProfServ      PctOccupManu      PctOccupMgmtProf      MalePctDivorce
## Min.   : 8.69    Min.   : 1.37    Min.   : 6.48    Min.   : 2.130
## 1st Qu.:20.07    1st Qu.: 9.13    1st Qu.:21.91    1st Qu.: 7.110
## Median :23.39    Median :13.15    Median :26.24    Median : 9.150
## Mean   :24.53    Mean   :13.82    Mean   :28.21    Mean   : 9.128
## 3rd Qu.:27.59    3rd Qu.:17.66    3rd Qu.:32.81    3rd Qu.:11.050
## Max.   :62.67    Max.   :44.27    Max.   :64.97    Max.   :20.080
##
##      MalePctNevMarr      FemalePctDiv      TotalPctDiv      PersPerFam
## Min.   :12.06    Min.   : 3.35    Min.   : 2.830    Min.   :2.29
## 1st Qu.:25.45    1st Qu.: 9.86    1st Qu.: 8.575    1st Qu.:2.99
## Median :29.00    Median :12.52    Median :10.900    Median :3.10
## Mean   :30.68    Mean   :12.33    Mean   :10.813    Mean   :3.13
## 3rd Qu.:33.41    3rd Qu.:14.74    3rd Qu.:12.985    3rd Qu.:3.22
## Max.   :76.60    Max.   :23.92    Max.   :22.230    Max.   :4.64
##
##      PctFam2Par      PctKids2Par      PctYoungKids2Par      PctTeen2Par
## Min.   :22.97    Min.   :18.30    Min.   : 8.70    Min.   :20.20
## 1st Qu.:67.90    1st Qu.:63.99    1st Qu.:74.78    1st Qu.:70.17
## Median :75.03    Median :72.53    Median :83.99    Median :76.92
## Mean   :74.06    Mean   :71.23    Mean   :81.87    Mean   :75.52

```

```

## 3rd Qu.:81.90 3rd Qu.:80.39 3rd Qu.: 91.67 3rd Qu.:82.77
## Max. :93.60 Max. :92.58 Max. :100.00 Max. :97.34
##
## PctWorkMomYoungKids PctWorkMom NumKidsBornNeverMar PctKidsBornNeverMar
## Min. :24.42 Min. :41.95 Min. : 0 Min. : 0.000
## 1st Qu.:55.43 1st Qu.:64.90 1st Qu.: 147 1st Qu.: 1.070
## Median :60.71 Median :69.23 Median : 352 Median : 2.040
## Mean :60.54 Mean :68.85 Mean : 2141 Mean : 3.115
## 3rd Qu.:65.98 3rd Qu.:73.50 3rd Qu.: 1032 3rd Qu.: 3.910
## Max. :87.97 Max. :89.37 Max. :527557 Max. :27.350
##
## NumImmig PctImmigRecent PctImmigRec5 PctImmigRec8
## Min. : 20 Min. : 0.000 Min. : 0.00 Min. : 0.00
## 1st Qu.: 400 1st Qu.: 6.695 1st Qu.:11.26 1st Qu.:17.20
## Median : 1024 Median :12.260 Median :19.08 Median :26.72
## Mean : 6277 Mean :13.526 Mean :20.42 Mean :27.54
## 3rd Qu.: 3302 3rd Qu.:17.950 3rd Qu.:27.45 3rd Qu.:36.49
## Max. :2082931 Max. :64.290 Max. :76.16 Max. :80.81
##
## PctImmigRec10 PctRecentImmig PctRecImmig5 PctRecImmig8
## Min. : 0.00 Min. : 0.000 Min. : 0.000 Min. : 0.000
## 1st Qu.:22.73 1st Qu.: 0.170 1st Qu.: 0.280 1st Qu.: 0.390
## Median :34.79 Median : 0.500 Median : 0.750 Median : 1.040
## Mean :34.73 Mean : 1.099 Mean : 1.697 Mean : 2.308
## 3rd Qu.:46.19 3rd Qu.: 1.310 3rd Qu.: 2.015 3rd Qu.: 2.700
## Max. :88.00 Max. :13.710 Max. :19.930 Max. :25.340
##
## PctRecImmig10 PctSpeakEnglOnly PctNotSpeakEnglWell PctLargHouseFam
## Min. : 0.000 Min. : 6.15 Min. : 0.000 Min. : 0.960
## 1st Qu.: 0.520 1st Qu.:84.38 1st Qu.: 0.510 1st Qu.: 3.390
## Median : 1.310 Median :92.18 Median : 0.920 Median : 4.280
## Mean : 2.944 Mean :87.07 Mean : 2.406 Mean : 5.387
## 3rd Qu.: 3.455 3rd Qu.:95.45 3rd Qu.: 2.270 3rd Qu.: 5.870
## Max. :32.630 Max. :98.98 Max. :38.330 Max. :34.870
##
## PctLargHouseOccup PersPerOccupHous PersPerOwnOccHous PersPerRentOccHous
## Min. : 0.440 Min. :1.580 Min. :1.61 Min. :1.550
## 1st Qu.: 2.370 1st Qu.:2.410 1st Qu.:2.55 1st Qu.:2.110
## Median : 3.050 Median :2.570 Median :2.71 Median :2.290
## Mean : 3.916 Mean :2.616 Mean :2.74 Mean :2.367
## 3rd Qu.: 4.210 3rd Qu.:2.770 3rd Qu.:2.90 3rd Qu.:2.530
## Max. :30.870 Max. :4.520 Max. :4.48 Max. :4.730
##
## PctPersOwnOccup PctPersDenseHous PctHousLess3BR MedNumBR
## Min. :13.93 Min. : 0.050 Min. : 3.06 Min. :1.000
## 1st Qu.:57.28 1st Qu.: 1.290 1st Qu.:37.51 1st Qu.:2.000
## Median :65.91 Median : 2.340 Median :46.39 Median :3.000
## Mean :66.37 Mean : 4.132 Mean :45.41 Mean :2.641
## 3rd Qu.:76.58 3rd Qu.: 4.730 3rd Qu.:53.52 3rd Qu.:3.000
## Max. :97.24 Max. :59.490 Max. :95.34 Max. :4.000
##
## HousVacant PctHousOccup PctHousOwnOcc PctVacantBoarded
## Min. : 36.0 Min. :37.47 Min. :16.86 Min. : 0.000
## 1st Qu.: 304.5 1st Qu.:91.29 1st Qu.:54.82 1st Qu.: 0.720

```

```

## Median : 558.0 Median :94.21 Median :62.83 Median : 1.660
## Mean : 1748.4 Mean :92.93 Mean :63.37 Mean : 2.779
## 3rd Qu.: 1228.0 3rd Qu.:96.02 3rd Qu.:72.64 3rd Qu.: 3.430
## Max. :172768.0 Max. :99.00 Max. :96.49 Max. :39.890
##
## PctVacMore6Mos MedYrHousBuilt PctHousNoPhone PctWOFullPlumb
## Min. : 3.12 Min. :1939 Min. : 0.000 Min. :0.0000
## 1st Qu.:24.48 1st Qu.:1956 1st Qu.: 0.905 1st Qu.:0.1600
## Median :34.10 Median :1964 Median : 2.850 Median :0.3200
## Mean :34.77 Mean :1963 Mean : 4.290 Mean :0.4253
## 3rd Qu.:43.97 3rd Qu.:1971 3rd Qu.: 6.805 3rd Qu.:0.5550
## Max. :82.13 Max. :1987 Max. :23.880 Max. :5.3300
##
## OwnOccLowQuart OwnOccMedVal OwnOccHiQuart OwnOccQrange
## Min. : 14999 Min. : 19500 Min. : 28200 Min. : 0
## 1st Qu.: 41500 1st Qu.: 56200 1st Qu.: 74300 1st Qu.: 32200
## Median : 65500 Median : 82800 Median :106700 Median : 43400
## Mean : 88696 Mean :113098 Mean :145318 Mean : 56623
## 3rd Qu.:121500 3rd Qu.:150600 3rd Qu.:188000 3rd Qu.: 65450
## Max. :500001 Max. :500001 Max. :500001 Max. :331000
##
## RentLowQ RentMedian RentHighQ RentQrange
## Min. : 99.0 Min. : 120.0 Min. : 182.0 Min. : 0.0
## 1st Qu.: 213.5 1st Qu.: 289.5 1st Qu.: 366.0 1st Qu.:139.0
## Median : 307.0 Median : 397.0 Median : 486.0 Median :171.0
## Mean : 330.0 Mean : 428.5 Mean : 527.3 Mean :197.3
## 3rd Qu.: 421.0 3rd Qu.: 544.0 3rd Qu.: 659.5 3rd Qu.:232.5
## Max. :1001.0 Max. :1001.0 Max. :1001.0 Max. :803.0
##
## MedRent MedRentPctHousInc MedOwnCostPctInc MedOwnCostPctIncNoMtg
## Min. : 192.0 Min. :14.9 Min. :14.00 Min. :10.10
## 1st Qu.: 364.0 1st Qu.:24.3 1st Qu.:18.70 1st Qu.:12.00
## Median : 467.0 Median :26.1 Median :21.00 Median :12.80
## Mean : 501.5 Mean :26.3 Mean :20.99 Mean :13.01
## 3rd Qu.: 615.0 3rd Qu.:28.0 3rd Qu.:23.10 3rd Qu.:13.70
## Max. :1001.0 Max. :35.1 Max. :32.70 Max. :23.40
##
## NumInShelters NumStreet PctForeignBorn PctBornSameState
## Min. : 0.00 Min. : 0.00 Min. : 0.18 Min. : 6.75
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 2.06 1st Qu.:50.11
## Median : 0.00 Median : 0.00 Median : 4.31 Median :64.49
## Mean : 66.95 Mean : 17.82 Mean : 7.34 Mean :61.54
## 3rd Qu.: 22.00 3rd Qu.: 1.00 3rd Qu.: 9.25 3rd Qu.:74.86
## Max. :23383.00 Max. :10447.00 Max. :60.40 Max. :93.14
##
## PctSameHouse85 PctSameCity85 PctSameState85 LemasSwornFT
## Min. :11.83 Min. :27.95 Min. :32.83 Min. : 65.0
## 1st Qu.:44.99 1st Qu.:72.06 1st Qu.:85.20 1st Qu.: 131.0
## Median :52.17 Median :79.49 Median :90.03 Median : 173.0
## Mean :51.54 Mean :77.41 Mean :88.11 Mean : 499.2
## 3rd Qu.:58.74 3rd Qu.:85.14 3rd Qu.:93.01 3rd Qu.: 314.0
## Max. :78.56 Max. :96.59 Max. :99.90 Max. :25655.0
## NA's :1872
## LemasSwFTPerPop LemasSwFTFieldOps LemasSwFTFieldPerPop LemasTotalReq

```


## Min. : 29.4	Min. : 14.0	Min. : 19.21	Min. : 2100
## 1st Qu.: 149.1	1st Qu.: 114.0	1st Qu.: 132.25	1st Qu.: 49864
## Median : 196.0	Median : 152.0	Median : 170.27	Median : 90000
## Mean : 246.5	Mean : 432.6	Mean : 210.84	Mean : 252405
## 3rd Qu.: 260.6	3rd Qu.: 285.5	3rd Qu.: 226.81	3rd Qu.: 171924
## Max. : 3437.2	Max. : 22496.0	Max. : 3290.62	Max. : 8328470
## NA's : 1872	NA's : 1872	NA's : 1872	NA's : 1872
## LemasTotReqPerPop	PolicReqPerOffic	PolicPerPop	RacialMatchCommPol
## Min. : 2705	Min. : 20.8	Min. : 29.4	Min. : 42.15
## 1st Qu.: 64845	1st Qu.: 343.4	1st Qu.: 149.2	1st Qu.: 79.44
## Median : 91035	Median : 443.2	Median : 196.0	Median : 87.93
## Mean : 120652	Mean : 523.7	Mean : 246.5	Mean : 85.50
## 3rd Qu.: 130325	3rd Qu.: 637.2	3rd Qu.: 260.6	3rd Qu.: 93.64
## Max. : 1926282	Max. : 2162.5	Max. : 3437.2	Max. : 100.00
## NA's : 1872	NA's : 1872	NA's : 1872	NA's : 1872
## PctPolicWhite	PctPolicBlack	PctPolicHisp	PctPolicAsian
## Min. : 1.60	Min. : 0.000	Min. : 0.000	Min. : 0.0000
## 1st Qu.: 76.23	1st Qu.: 2.020	1st Qu.: 0.415	1st Qu.: 0.0000
## Median : 86.18	Median : 5.000	Median : 2.040	Median : 0.0000
## Mean : 82.52	Mean : 9.263	Mean : 5.460	Mean : 0.6813
## 3rd Qu.: 93.34	3rd Qu.: 14.065	3rd Qu.: 6.215	3rd Qu.: 0.6500
## Max. : 100.00	Max. : 67.310	Max. : 98.400	Max. : 18.5700
## NA's : 1872	NA's : 1872	NA's : 1872	NA's : 1872
## PctPolicMinor	OfficAssgnDrugUnits	NumKindsDrugsSeiz	PolicAveOTWorked
## Min. : 0.00	Min. : 0.00	Min. : 1.000	Min. : 0.0
## 1st Qu.: 4.95	1st Qu.: 6.00	1st Qu.: 7.000	1st Qu.: 55.1
## Median : 11.37	Median : 12.00	Median : 9.000	Median : 98.7
## Mean : 15.24	Mean : 26.29	Mean : 8.816	Mean : 119.1
## 3rd Qu.: 19.74	3rd Qu.: 23.00	3rd Qu.: 10.500	3rd Qu.: 153.6
## Max. : 98.40	Max. : 1773.00	Max. : 15.000	Max. : 634.7
## NA's : 1872	NA's : 1872	NA's : 1872	NA's : 1872
## LandArea	PopDens	PctUsePubTrans	PolicCars
## Min. : 0.90	Min. : 10	Min. : 0.000	Min. : 20.0
## 1st Qu.: 7.30	1st Qu.: 1182	1st Qu.: 0.360	1st Qu.: 54.0
## Median : 13.70	Median : 2027	Median : 1.220	Median : 86.0
## Mean : 27.42	Mean : 2784	Mean : 3.041	Mean : 185.5
## 3rd Qu.: 26.10	3rd Qu.: 3322	3rd Qu.: 3.365	3rd Qu.: 189.5
## Max. : 3569.80	Max. : 44230	Max. : 54.330	Max. : 3187.0
##			NA's : 1872
## PolicOperBudg	LemasPctPolicOnPatr	LemasGangUnitDeploy	
## Min. : 2.380e+06	Min. : 10.85	Min. : 0.000	
## 1st Qu.: 7.275e+06	1st Qu.: 84.30	1st Qu.: 0.000	
## Median : 1.116e+07	Median : 89.58	Median : 5.000	
## Mean : 3.218e+07	Mean : 87.13	Mean : 4.286	
## 3rd Qu.: 2.015e+07	3rd Qu.: 93.20	3rd Qu.: 10.000	
## Max. : 1.617e+09	Max. : 99.94	Max. : 10.000	
## NA's : 1872	NA's : 1872	NA's : 1872	
## LemasPctOfficDrugUn	PolicBudgPerPop	murders	murderPerPop
## Min. : 0.0000	Min. : 15260	Min. : 0.000	Min. : 0.000
## 1st Qu.: 0.0000	1st Qu.: 88094	1st Qu.: 0.000	1st Qu.: 0.000
## Median : 0.0000	Median : 114582	Median : 1.000	Median : 2.170
## Mean : 0.9802	Mean : 153578	Mean : 7.765	Mean : 5.859
## 3rd Qu.: 0.0000	3rd Qu.: 155656	3rd Qu.: 3.000	3rd Qu.: 8.365
## Max. : 48.4400	Max. : 2422367	Max. : 1946.000	Max. : 91.090

```

##          NA's :1872
##      rapes      rapesPerPop      robberies      robbbPerPop
## Min.   : 0.00   Min.   : 0.00   Min.   : 0.0   Min.   : 0.00
## 1st Qu.: 2.00   1st Qu.: 11.54   1st Qu.: 5.0   1st Qu.: 27.65
## Median : 7.00   Median : 26.92   Median : 19.0   Median : 74.80
## Mean   : 28.05   Mean   : 36.26   Mean   : 237.9   Mean   : 162.61
## 3rd Qu.: 19.00   3rd Qu.: 51.47   3rd Qu.: 70.0   3rd Qu.: 187.16
## Max.   :2818.00   Max.   :401.35   Max.   :86001.0   Max.   :2264.13
## NA's   :208     NA's   :208     NA's    :1       NA's    :1
##      assaults      assaultPerPop      burglaries      burglPerPop
## Min.   : 0.0   Min.   : 0.00   Min.   : 2.0   Min.   : 16.92
## 1st Qu.: 18.0   1st Qu.: 94.19   1st Qu.: 95.0   1st Qu.: 511.69
## Median : 56.0   Median : 226.53   Median : 205.0   Median : 822.72
## Mean   : 326.5   Mean   : 378.00   Mean   : 761.2   Mean   : 1033.43
## 3rd Qu.: 180.0   3rd Qu.: 504.39   3rd Qu.: 508.0   3rd Qu.: 1350.23
## Max.   :62778.0   Max.   :4932.50   Max.   :99207.0   Max.   :11881.02
## NA's   :13      NA's   :13      NA's    :3       NA's    :3
##      larcenies      larcPerPop      autoTheft      autoTheftPerPop
## Min.   : 10   Min.   : 77.86   Min.   : 1.0   Min.   : 6.55
## 1st Qu.: 392   1st Qu.: 2040.08   1st Qu.: 30.0   1st Qu.: 156.95
## Median : 747   Median : 3079.51   Median : 75.0   Median : 302.36
## Mean   : 2138   Mean   : 3372.98   Mean   : 516.7   Mean   : 473.97
## 3rd Qu.: 1675   3rd Qu.: 4335.41   3rd Qu.: 232.5   3rd Qu.: 589.77
## Max.   :235132   Max.   :25910.55   Max.   :112464.0   Max.   :4968.59
## NA's   :3       NA's   :3       NA's    :3       NA's    :3
##      arsons      arsonsPerPop      ViolentCrimesPerPop      nonViolPerPop
## Min.   : 0.00   Min.   : 0.00   Min.   : 0.0   Min.   : 116.8
## 1st Qu.: 1.00   1st Qu.: 7.67   1st Qu.: 161.7   1st Qu.: 2918.1
## Median : 5.00   Median : 21.08   Median : 374.1   Median : 4425.4
## Mean   : 30.91   Mean   : 32.15   Mean   : 589.1   Mean   : 4908.2
## 3rd Qu.: 16.00   3rd Qu.: 42.85   3rd Qu.: 794.4   3rd Qu.: 6229.3
## Max.   :5119.00   Max.   :436.37   Max.   :4877.1   Max.   :27119.8
## NA's   :91      NA's   :91      NA's    :221     NA's    :97

```

```
str(crime)
```

```

## 'data.frame': 2215 obs. of 147 variables:
## $ Êcommunityname : chr "BerkeleyHeightstownship" "Marpletownship" "Tigardcity" "Gloversville"
## $ state : chr "NJ" "PA" "OR" "NY" ...
## $ countyCode : int 39 45 NA 35 7 NA 21 NA 17 NA ...
## $ communityCode : int 5320 47616 NA 29443 5068 NA 50250 NA 25700 NA ...
## $ fold : int 1 1 1 1 1 1 1 1 1 1 ...
## $ population : int 11980 23123 29344 16656 11245 140494 28700 59459 74111 103590 ...
## $ householdsize : num 3.1 2.82 2.43 2.4 2.76 2.45 2.6 2.45 2.46 2.62 ...
## $ racePctblack : num 1.37 0.8 0.74 1.7 0.53 ...
## $ racePctWhite : num 91.8 95.6 94.3 97.3 89.2 ...
## $ racePctAsian : num 6.5 3.44 3.43 0.5 1.17 0.9 1.47 0.4 1.25 0.92 ...
## $ racePctHisp : num 1.88 0.85 2.35 0.7 0.52 ...
## $ agePct12t21 : num 12.5 11 11.4 12.6 24.5 ...
## $ agePct12t29 : num 21.4 21.3 25.9 25.2 40.5 ...
## $ agePct16t24 : num 10.9 10.5 11 12.2 28.7 ...
## $ agePct65up : num 11.3 17.2 10.3 17.6 12.6 ...
## $ numbUrban : int 11980 23123 29344 0 0 140494 28700 59449 74115 103590 ...
## $ pctUrban : num 100 100 100 0 0 100 100 100 100 100 ...

```

```

## $ medIncome      : int 75122 47917 35669 20580 17390 21577 42805 23221 25326 17852 ...
## $ pctWWage       : num 89.2 79 82 68.2 69.3 ...
## $ pctWFarmSelf   : num 1.55 1.11 1.15 0.24 0.55 1 0.39 0.67 2.93 0.86 ...
## $ pctWInvInc     : num 70.2 64.1 55.7 39 42.8 ...
## $ pctWSocSec     : num 23.6 35.5 22.2 39.5 32.2 ...
## $ pctWPubAsst    : num 1.03 2.75 2.94 11.71 11.21 ...
## $ pctWRetire     : num 18.4 22.9 14.6 18.3 14.4 ...
## $ medFamInc      : int 79584 55323 42112 26501 24018 27705 50394 28901 34269 24058 ...
## $ perCapInc       : int 29711 20148 16946 10810 8483 11878 18193 12161 13554 10195 ...
## $ whitePerCap     : int 30233 20191 17103 10909 9009 12029 18276 12599 13727 12126 ...
## $ blackPerCap     : int 13600 18137 16644 9984 887 7382 17342 9820 8852 5715 ...
## $ indianPerCap    : int 5725 0 21606 4941 4425 10264 21482 6634 5344 11313 ...
## $ AsianPerCap     : int 27101 20074 15528 3541 3352 10753 12639 8802 8011 5770 ...
## $ OtherPerCap     : int 5115 5250 5954 2451 3000 7192 21852 7428 5332 7320 ...
## $ HispPerCap      : int 22838 12222 8405 4391 1328 8104 22594 6187 5174 6984 ...
## $ NumUnderPov     : int 227 885 1389 2831 2855 23223 1126 10320 9603 27767 ...
## $ PctPopUnderPov  : num 1.96 3.98 4.75 17.23 29.99 ...
## $ PctLess9thGrade : num 5.81 5.61 2.8 11.05 12.15 ...
## $ PctNotHSGrad     : num 9.9 13.72 9.09 33.68 23.06 ...
## $ PctBSorMore      : num 48.2 29.9 30.1 10.8 25.3 ...
## $ PctUnemployed    : num 2.7 2.43 4.01 9.86 9.08 5.72 4.85 8.19 4.18 8.39 ...
## $ PctEmploy        : num 64.5 62 69.8 54.7 52.4 ...
## $ PctEmplManu      : num 14.65 12.26 15.95 31.22 6.89 ...
## $ PctEmplProfServ  : num 28.8 29.3 21.5 27.4 36.5 ...
## $ PctOccupManu     : num 5.49 6.39 8.79 26.76 10.94 ...
## $ PctOccupMgmtProf : num 50.7 37.6 32.5 22.7 27.8 ...
## $ MalePctDivorce   : num 3.67 4.23 10.1 10.98 7.51 ...
## $ MalePctNevMarr   : num 26.4 28 25.8 28.1 50.7 ...
## $ FemalePctDiv     : num 5.22 6.45 14.76 14.47 11.64 ...
## $ TotalPctDiv      : num 4.47 5.42 12.55 12.91 9.73 ...
## $ PersPerFam       : num 3.22 3.11 2.95 2.98 2.98 2.89 3.14 2.95 3 3.11 ...
## $ PctFam2Par       : num 91.4 86.9 78.5 64 58.6 ...
## $ PctKids2Par      : num 90.2 85.3 78.8 62.4 55.2 ...
## $ PctYoungKids2Par : num 95.8 96.8 92.4 65.4 66.5 ...
## $ PctTeen2Par      : num 95.8 86.5 75.7 67.4 79.2 ...
## $ PctWorkMomYoungKids : num 44.6 51.1 66.1 59.6 61.2 ...
## $ PctWorkMom       : num 58.9 62.4 74.2 70.3 68.9 ...
## $ NumKidsBornNeverMar : int 31 43 164 561 402 1511 263 2368 751 3537 ...
## $ PctKidsBornNeverMar : num 0.36 0.24 0.88 3.84 4.7 1.58 1.18 4.66 1.64 4.71 ...
## $ NumImmig         : int 1277 1920 1468 339 196 2091 2637 517 1474 4793 ...
## $ PctImmigRecent   : num 8.69 5.21 16.42 13.86 46.94 ...
## $ PctImmigRec5     : num 13 8.65 23.98 13.86 56.12 ...
## $ PctImmigRec8     : num 21 13.3 32.1 15.3 67.9 ...
## $ PctImmigRec10    : num 30.9 22.5 35.6 15.3 69.9 ...
## $ PctRecentImmig   : num 0.93 0.43 0.82 0.28 0.82 0.32 1.05 0.11 0.47 0.72 ...
## $ PctRecImmig5     : num 1.39 0.72 1.2 0.28 0.98 0.45 1.49 0.2 0.67 1.07 ...
## $ PctRecImmig8     : num 2.24 1.11 1.61 0.31 1.18 0.57 2.2 0.25 0.93 1.63 ...
## $ PctRecImmig10    : num 3.3 1.87 1.78 0.31 1.22 0.68 2.55 0.29 1.07 2.31 ...
## $ PctSpeakEnglOnly : num 85.7 87.8 93.1 95 94.6 ...
## $ PctNotSpeakEnglWell : num 1.37 1.81 1.14 0.56 0.39 0.6 0.6 0.28 0.43 2.51 ...
## $ PctLargHouseFam  : num 4.81 4.25 2.97 3.93 5.23 3.08 5.08 3.85 2.59 6.7 ...
## $ PctLargHouseOccup : num 4.17 3.34 2.05 2.56 3.11 1.92 3.46 2.55 1.54 4.1 ...
## $ PersPerOccupHous : num 2.99 2.7 2.42 2.37 2.35 2.28 2.55 2.36 2.32 2.45 ...
## $ PersPerOwnOccHous : num 3 2.83 2.69 2.51 2.55 2.37 2.89 2.42 2.77 2.47 ...

```

```
## $ PersPerRentOccHous : num 2.84 1.96 2.06 2.2 2.12 2.16 2.09 2.27 1.91 2.44 ...
## $ PctPersOwnOccup : num 91.5 89 64.2 58.2 58.1 ...
## $ PctPersDenseHous : num 0.39 1.01 2.03 1.21 2.94 2.11 1.47 1.9 1.67 6.14 ...
## $ PctHousLess3BR : num 11.1 23.6 47.5 45.7 55.6 ...
## $ MedNumBR : int 3 3 3 3 2 2 3 2 2 2 ...
## $ HousVacant : int 64 240 544 669 333 5119 566 2051 1562 5606 ...
## $ PctHousOccup : num 98.4 97.2 95.7 91.2 92.5 ...
## $ PctHousOwnOcc : num 91 84.9 57.8 54.9 53.6 ...
## $ PctVacantBoarded : num 3.12 0 0.92 2.54 3.9 2.09 1.41 6.39 0.45 5.64 ...
## $ PctVacMore6Mos : num 37.5 18.33 7.54 57.85 42.64 ...
## $ MedYrHousBuilt : int 1959 1958 1976 1939 1958 1966 1956 1954 1971 1960 ...
## $ PctHousNoPhone : num 0 0.31 1.55 7 7.45 ...
## $ PctWOFullPlumb : num 0.28 0.14 0.12 0.87 0.82 0.31 0.28 0.49 0.19 0.33 ...
## $ OwnOccLowQuart : int 215900 136300 74700 36400 30600 37700 155100 26300 54500 28600 ...
## $ OwnOccMedVal : int 262600 164200 90400 49600 43200 53900 179000 37000 70300 43100 ...
## $ OwnOccHiQuart : int 326900 199900 112000 66500 59500 73100 215500 52400 93700 67400 ...
## $ OwnOccQrange : int 111000 63600 37300 30100 28900 35400 60400 26100 39200 38800 ...
## $ RentLowQ : int 685 467 370 195 202 215 463 186 241 192 ...
## $ RentMedian : int 1001 560 428 250 283 280 669 253 321 281 ...
## $ RentHighQ : int 1001 672 520 309 362 349 824 325 387 369 ...
## $ RentQrange : int 316 205 150 114 160 134 361 139 146 177 ...
## $ MedRent : int 1001 627 484 333 332 340 736 338 355 353 ...
## $ MedRentPctHousInc : num 23.8 27.6 24.1 28.7 32.2 26.4 24.4 26.3 25.2 29.6 ...
## $ MedOwnCostPctInc : num 21.1 20.7 21.7 20.6 23.2 17.3 20.8 15.1 20.7 19.4 ...
## $ MedOwnCostPctIncNoMtg : num 14 12.5 11.6 14.5 12.9 11.7 12.5 12.2 12.8 13 ...
## $ NumInShelters : int 11 0 16 0 2 327 0 21 125 43 ...
## $ NumStreet : int 0 0 0 0 0 4 0 0 15 4 ...
## $ PctForeignBorn : num 10.66 8.3 5 2.04 1.74 ...
## [list output truncated]
```

```
sum(is.na(crime) == TRUE)
```

```
## [1] 44592
```

```
# This step shows there are 44592 missing values
# in total.
```

```
sort(sapply(crime, function(x) sum(is.na(x))))
```

```
##      Ecommunityname      state      fold
##      0      0      0
##      population      householdsize      racepctblack
##      0      0      0
##      racePctWhite      racePctAsian      racePctHisp
##      0      0      0
##      agePct12t21      agePct12t29      agePct16t24
##      0      0      0
##      agePct65up      numbUrban      pctUrban
##      0      0      0
##      medIncome      pctWWage      pctWFarmSelf
##      0      0      0
##      pctWInvInc      pctWSocSec      pctWPubAsst
##      0      0      0
```

##	pctWRetire	medFamInc	perCapInc
##	0	0	0
##	whitePerCap	blackPerCap	indianPerCap
##	0	0	0
##	AsianPerCap	HispPerCap	NumUnderPov
##	0	0	0
##	PctPopUnderPov	PctLess9thGrade	PctNotHSGrad
##	0	0	0
##	PctBSorMore	PctUnemployed	PctEmploy
##	0	0	0
##	PctEmplManu	PctEmplProfServ	PctOccupManu
##	0	0	0
##	PctOccupMgmtProf	MalePctDivorce	MalePctNevMarr
##	0	0	0
##	FemalePctDiv	TotalPctDiv	PersPerFam
##	0	0	0
##	PctFam2Par	PctKids2Par	PctYoungKids2Par
##	0	0	0
##	PctTeen2Par	PctWorkMomYoungKids	PctWorkMom
##	0	0	0
##	NumKidsBornNeverMar	PctKidsBornNeverMar	NumImmig
##	0	0	0
##	PctImmigRecent	PctImmigRec5	PctImmigRec8
##	0	0	0
##	PctImmigRec10	PctRecentImmig	PctRecImmig5
##	0	0	0
##	PctRecImmig8	PctRecImmig10	PctSpeakEnglOnly
##	0	0	0
##	PctNotSpeakEnglWell	PctLargHouseFam	PctLargHouseOccup
##	0	0	0
##	PersPerOccupHous	PersPerOwnOccHous	PersPerRentOccHous
##	0	0	0
##	PctPersOwnOccup	PctPersDenseHous	PctHousLess3BR
##	0	0	0
##	MedNumBR	HousVacant	PctHousOccup
##	0	0	0
##	PctHousOwnOcc	PctVacantBoarded	PctVacMore6Mos
##	0	0	0
##	MedYrHousBuilt	PctHousNoPhone	PctWOFullPlumb
##	0	0	0
##	OwnOccLowQuart	OwnOccMedVal	OwnOccHiQuart
##	0	0	0
##	OwnOccQrange	RentLowQ	RentMedian
##	0	0	0
##	RentHighQ	RentQrange	MedRent
##	0	0	0
##	MedRentPctHousInc	MedOwnCostPctInc	MedOwnCostPctIncNoMtg
##	0	0	0
##	NumInShelters	NumStreet	PctForeignBorn
##	0	0	0
##	PctBornSameState	PctSameHouse85	PctSameCity85
##	0	0	0
##	PctSameState85	LandArea	PopDens
##	0	0	0

```
##      PctUsePubTrans  LemasPctOfficDrugUn      murders
##              0              0              0
##      murdPerPop      OtherPerCap      robberies
##              0              1              1
##      robbbPerPop      burglaries      burglPerPop
##              1              3              3
##      larcenies      larcPerPop      autoTheft
##              3              3              3
##      autoTheftPerPop      assaults      assaultPerPop
##              3              13             13
##      arsons      arsonsPerPop      nonViolPerPop
##              91              91             97
##      rapes      rapesPerPop      ViolentCrimesPerPop
##              208             208             221
##      countyCode      communityCode      LemasSwornFT
##              1221             1224             1872
##      LemasSwFTPerPop      LemasSwFTFieldOps      LemasSwFTFieldPerPop
##              1872             1872             1872
##      LemasTotalReq      LemasTotReqPerPop      PolicReqPerOffic
##              1872             1872             1872
##      PolicPerPop      RacialMatchCommPol      PctPolicWhite
##              1872             1872             1872
##      PctPolicBlack      PctPolicHisp      PctPolicAsian
##              1872             1872             1872
##      PctPolicMinor      OfficAssgnDrugUnits      NumKindsDrugsSeiz
##              1872             1872             1872
##      PolicAveOTWorked      PolicCars      PolicOperBudg
##              1872             1872             1872
##      LemasPctPolicOnPatr      LemasGangUnitDeploy      PolicBudgPerPop
##              1872             1872             1872
```

```
# This step provides a list of NA value for each
# variable, ordered by ascending order for the
# number of NAs.
```

```
table(complete.cases(crime))
```

```
##
## FALSE  TRUE
## 2104   111
```

```
prop.table(table(complete.cases(crime))) * 100
```

```
##
##      FALSE      TRUE
## 94.988713  5.011287
```

```
# The complete cases ratio is only around 5% in
# this data set. Approximately 95% entries have
# missing data. This is not surprising because
# many variables have up to 1872 missing values
# out of 2215 total observations.
```

```
library(devtools)
```

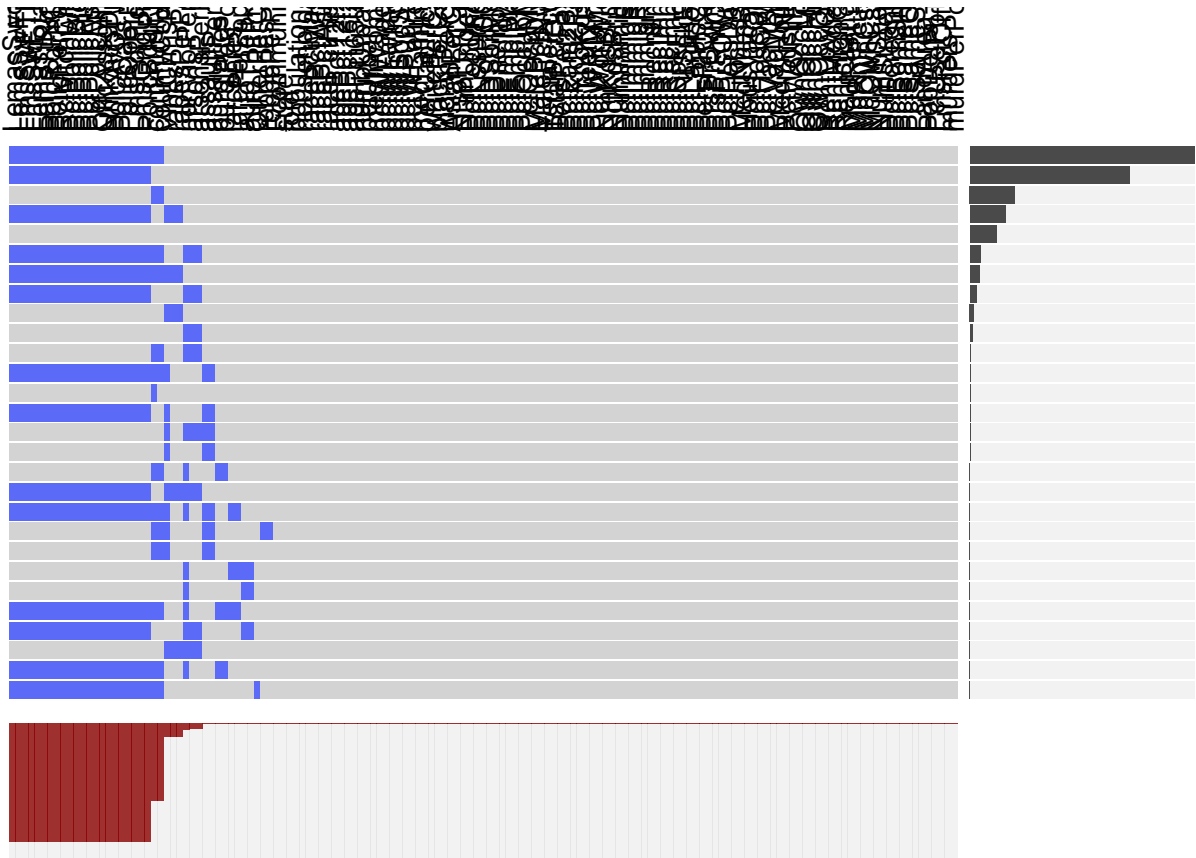
```
## Loading required package: usethis
```

```
library(extracat)
```

```
visna(crime, sort = "b")
```

```
## Warning in melt(as.data.frame(xs), ncol(xs)): The melt generic in data.table  
## has been passed a data.frame and will attempt to redirect to the relevant  
## reshape2 method; please note that reshape2 is deprecated, and this redirection  
## is now deprecated as well. To continue using melt methods from reshape2 while  
## both libraries are attached, e.g. melt.list, you can prepend the namespace like  
## reshape2::melt(as.data.frame(xs)). In the next version, this warning will become  
## an error.
```

```
## Warning in abbreviate(dimnames(tab)[[2]], abbrev[2]): abbreviate used with non-  
## ASCII chars
```



```
# The visualization shows that lots of missing  
# values are concentrated on several variables.  
# The range of over 50% data missing starts from  
# countrycode (1221 missing out of 2215) to
```

```
# PolicBudgPerPop (1872 missing out of 2215). The
# range includes 24 variables out of 147
# variables in total. Other 123 variables have
# less than 222 missing values. The rule of
# thumb is to delete the 24 variables with over
# 50% missing values first.
```

```
# Defining & Treating missing values The missing
# values follow the MCAR features because the
# missing pattern doesn't depend on any variable.
# They are missing completely at random. Since
# the missing values are over 50% for 24
# variables, column-wise deletion will be
# executed firstly.
```

```
colnames(crime)
```

```
##      [1] "Ecommunityname"      "state"          "countyCode"
##      [4] "communityCode"      "fold"           "population"
##      [7] "householdsize"      "racepctblack"   "racePctWhite"
##     [10] "racePctAsian"       "racePctHisp"    "agePct12t21"
##     [13] "agePct12t29"        "agePct16t24"    "agePct65up"
##     [16] "numbUrban"          "pctUrban"       "medIncome"
##     [19] "pctWWage"           "pctWFarmSelf"   "pctWInvInc"
##     [22] "pctWSocSec"         "pctWPubAsst"    "pctWRetire"
##     [25] "medFamInc"          "perCapInc"      "whitePerCap"
##     [28] "blackPerCap"        "indianPerCap"   "AsianPerCap"
##     [31] "OtherPerCap"        "HispPerCap"     "NumUnderPov"
##     [34] "PctPopUnderPov"     "PctLess9thGrade" "PctNotHSGrad"
##     [37] "PctBSorMore"        "PctUnemployed"  "PctEmploy"
##     [40] "PctEmplManu"        "PctEmplProfServ" "PctOccupManu"
##     [43] "PctOccupMgmtProf"   "MalePctDivorce" "MalePctNevMarr"
##     [46] "FemalePctDiv"       "TotalPctDiv"    "PersPerFam"
##     [49] "PctFam2Par"         "PctKids2Par"    "PctYoungKids2Par"
##     [52] "PctTeen2Par"        "PctWorkMomYoungKids" "PctWorkMom"
##     [55] "NumKidsBornNeverMar" "PctKidsBornNeverMar" "NumImmig"
##     [58] "PctImmigRecent"     "PctImmigRec5"   "PctImmigRec8"
##     [61] "PctImmigRec10"      "PctRecentImmig" "PctRecImmig5"
##     [64] "PctRecImmig8"       "PctRecImmig10"  "PctSpeakEnglOnly"
##     [67] "PctNotSpeakEnglWell" "PctLargHouseFam" "PctLargHouseOccup"
##     [70] "PersPerOccupHous"   "PersPerOwnOccHous" "PersPerRentOccHous"
##     [73] "PctPersOwnOccup"    "PctPersDenseHous" "PctHousLess3BR"
##     [76] "MedNumBR"           "HousVacant"     "PctHousOccup"
##     [79] "PctHousOwnOcc"      "PctVacantBoarded" "PctVacMore6Mos"
##     [82] "MedYrHousBuilt"     "PctHousNoPhone" "PctWOFullPlumb"
##     [85] "OwnOccLowQuart"     "OwnOccMedVal"   "OwnOccHiQuart"
##     [88] "OwnOccQrange"       "RentLowQ"       "RentMedian"
##     [91] "RentHighQ"          "RentQrange"     "MedRent"
##     [94] "MedRentPctHousInc"  "MedOwnCostPctInc" "MedOwnCostPctIncNoMtg"
##     [97] "NumInShelters"      "NumStreet"      "PctForeignBorn"
##    [100] "PctBornSameState"   "PctSameHouse85" "PctSameCity85"
##    [103] "PctSameState85"     "LemasSwornFT"   "LemasSwFTPerPop"
```



```
## [106] "LemasSwFTFieldOps"      "LemasSwFTFieldPerPop"  "LemasTotalReq"
## [109] "LemasTotReqPerPop"      "PolicReqPerOffic"      "PolicPerPop"
## [112] "RacialMatchCommPol"     "PctPolicWhite"         "PctPolicBlack"
## [115] "PctPolicHisp"           "PctPolicAsian"         "PctPolicMinor"
## [118] "OfficAssgnDrugUnits"    "NumKindsDrugsSeiz"     "PolicAveOTWorked"
## [121] "LandArea"               "PopDens"               "PctUsePubTrans"
## [124] "PolicCars"              "PolicOperBudg"         "LemasPctPolicOnPatr"
## [127] "LemasGangUnitDeploy"    "LemasPctOfficDrugUn"   "PolicBudgPerPop"
## [130] "murders"                "murdPerPop"            "rapes"
## [133] "rapesPerPop"            "robberies"              "robbbPerPop"
## [136] "assaults"               "assaultPerPop"         "burglaries"
## [139] "burglPerPop"            "larcenies"              "larcPerPop"
## [142] "autoTheft"              "autoTheftPerPop"       "arsons"
## [145] "arsonsPerPop"           "ViolentCrimesPerPop"   "nonViolPerPop"
```

```
crime2 <- crime[-c(3, 4, 104:120, 124:128)]
```

```
# For the variables with 1-221 missing values,
# mean imputation would be a good way to preserve
# the records while not misleading the outcome of
# analysis. To do this I use the following codes.
library(Hmisc)
```

```
## Warning: package 'Hmisc' was built under R version 4.0.5
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      format.pval, units
```

```
with(crime2, impute(robberies, mean))
with(crime2, impute(robbbPerPop, mean))
with(crime2, impute(burglaries, mean))
with(crime2, impute(burglPerPop, mean))
with(crime2, impute(larcenies, mean))
with(crime2, impute(larcPerPop, mean))
with(crime2, impute(assaults, mean))
with(crime2, impute(assaultPerPop, mean))
with(crime2, impute(arsons, mean))
with(crime2, impute(arsonsPerPop, mean))
with(crime2, impute(nonViolPerPop, mean))
with(crime2, impute(rapes, mean))
with(crime2, impute(rapesPerPop, mean))
with(crime2, impute(ViolentCrimesPerPop, mean))
```

```
# Now let us look at the updated data summary.
# In short, the original data is preserved in an
# optimal way without any record deletion. The
# only deletion occurred for the 24 variables
# with over 50% values missing across records.
# They were excluded from the table. However, if
# there was the need to highlight the heavily
# incomplete data, record deletion or pair-wise
# deletion might be more appropriate. That would
# lead to a smaller-scale analysis focusing on
# those variables instead of the holistic
# analysis.
```

```
summary(crime2)
```

```
##      Ecommunityname      state      fold      population
## Length:2215      Length:2215      Min.   : 1.000      Min.   : 10005
## Class :character  Class :character  1st Qu.: 3.000      1st Qu.: 14366
## Mode  :character  Mode  :character  Median : 5.000      Median : 22792
##                                     Mean  : 5.494      Mean  : 53118
##                                     3rd Qu.: 8.000      3rd Qu.: 43024
##                                     Max.   :10.000      Max.   :7322564
##
## householdsize      racepctblack      racePctWhite      racePctAsian
## Min.   :1.600      Min.   : 0.000      Min.   : 2.68      Min.   : 0.03
## 1st Qu.:2.500      1st Qu.: 0.860      1st Qu.:76.32      1st Qu.: 0.62
## Median :2.660      Median : 2.870      Median :90.35      Median : 1.23
## Mean   :2.707      Mean   : 9.335      Mean   :83.98      Mean   : 2.67
## 3rd Qu.:2.850      3rd Qu.:11.145      3rd Qu.:96.22      3rd Qu.: 2.67
## Max.   :5.280      Max.   :96.670      Max.   :99.63      Max.   :57.46
##
##      racePctHispanic      agePct12t21      agePct12t29      agePct16t24
## Min.   : 0.12      Min.   : 4.58      Min.   : 9.38      Min.   : 4.64
## 1st Qu.: 0.93      1st Qu.:12.25      1st Qu.:24.41      1st Qu.:11.32
## Median : 2.18      Median :13.62      Median :26.78      Median :12.54
## Mean   : 7.95      Mean   :14.45      Mean   :27.64      Mean   :13.98
## 3rd Qu.: 7.81      3rd Qu.:15.36      3rd Qu.:29.20      3rd Qu.:14.35
## Max.   :95.29      Max.   :54.40      Max.   :70.51      Max.   :63.62
##
##      agePct65up      numbUrban      pctUrban      medIncome
## Min.   : 1.66      Min.   : 0      Min.   : 0.00      Min.   : 8866
## 1st Qu.: 8.75      1st Qu.: 0      1st Qu.: 0.00      1st Qu.: 23817
## Median :11.73      Median : 18041      Median :100.00      Median : 31441
## Mean   :11.84      Mean   : 47735      Mean   : 70.47      Mean   : 33985
## 3rd Qu.:14.41      3rd Qu.: 41918      3rd Qu.:100.00      3rd Qu.: 41481
## Max.   :52.77      Max.   :7322564      Max.   :100.00      Max.   :123625
##
##      pctWWage      pctWFarmSelf      pctWInvInc      pctWSocSec
## Min.   :31.68      Min.   :0.0000      Min.   : 5.81      Min.   : 4.81
## 1st Qu.:73.40      1st Qu.:0.4600      1st Qu.:34.68      1st Qu.:20.77
## Median :78.61      Median :0.6900      Median :42.88      Median :26.59
## Mean   :78.31      Mean   :0.8818      Mean   :43.75      Mean   :26.41
## 3rd Qu.:84.03      3rd Qu.:1.1000      3rd Qu.:52.74      3rd Qu.:31.71
```

```

## Max. :96.76 Max. :6.5300 Max. :89.04 Max. :76.39
##
## pctWPubAsst pctWRetire medFamInc perCapInc
## Min. : 0.180 Min. : 3.46 Min. : 10447 Min. : 5237
## 1st Qu.: 3.270 1st Qu.:12.88 1st Qu.: 29538 1st Qu.:11602
## Median : 5.610 Median :15.65 Median : 36678 Median :14101
## Mean : 6.801 Mean :15.97 Mean : 39857 Mean :15604
## 3rd Qu.: 9.105 3rd Qu.:18.75 3rd Qu.: 46999 3rd Qu.:17795
## Max. :44.820 Max. :45.51 Max. :139008 Max. :63302
##
## whitePerCap blackPerCap indianPerCap AsianPerCap
## Min. : 5472 Min. : 0 Min. : 0 Min. : 0
## 1st Qu.:12610 1st Qu.: 6742 1st Qu.: 6345 1st Qu.: 8286
## Median :15073 Median : 9777 Median : 9895 Median : 12250
## Mean :16568 Mean : 11542 Mean : 12229 Mean : 14228
## 3rd Qu.:18610 3rd Qu.:14526 3rd Qu.:14758 3rd Qu.:17328
## Max. :68850 Max. :212120 Max. :480000 Max. :106165
##
## OtherPerCap HispPerCap NumUnderPov PctPopUnderPov
## Min. : 0 Min. : 0 Min. : 78.0 Min. : 0.64
## 1st Qu.: 5528 1st Qu.: 7274 1st Qu.: 912.5 1st Qu.: 4.51
## Median : 8186 Median : 9721 Median : 2142.0 Median : 9.33
## Mean : 9443 Mean :11019 Mean : 7590.9 Mean :11.62
## 3rd Qu.:11526 3rd Qu.:13418 3rd Qu.: 4988.0 3rd Qu.:16.91
## Max. :137000 Max. :54648 Max. :1384994.0 Max. :58.00
## NA's :1
## PctLess9thGrade PctNotHSGrad PctBSorMore PctUnemployed
## Min. : 0.200 Min. : 1.46 Min. : 1.63 Min. : 1.320
## 1st Qu.: 4.640 1st Qu.:13.92 1st Qu.:14.10 1st Qu.: 4.045
## Median : 7.740 Median :21.38 Median :19.65 Median : 5.450
## Mean : 9.187 Mean :22.31 Mean :23.06 Mean : 6.045
## 3rd Qu.:11.835 3rd Qu.:29.20 3rd Qu.:29.05 3rd Qu.: 7.440
## Max. :49.890 Max. :73.66 Max. :79.18 Max. :31.230
##
## PctEmploy PctEmplManu PctEmplProfServ PctOccupManu
## Min. :24.82 Min. : 2.05 Min. : 8.69 Min. : 1.37
## 1st Qu.:56.49 1st Qu.:12.21 1st Qu.:20.07 1st Qu.: 9.13
## Median :62.44 Median :17.30 Median :23.39 Median :13.15
## Mean :62.02 Mean :18.23 Mean :24.53 Mean :13.82
## 3rd Qu.:67.83 3rd Qu.:23.40 3rd Qu.:27.59 3rd Qu.:17.66
## Max. :84.67 Max. :50.03 Max. :62.67 Max. :44.27
##
## PctOccupMgmtProf MalePctDivorce MalePctNevMarr FemalePctDiv
## Min. : 6.48 Min. : 2.130 Min. :12.06 Min. : 3.35
## 1st Qu.:21.91 1st Qu.: 7.110 1st Qu.:25.45 1st Qu.: 9.86
## Median :26.24 Median : 9.150 Median :29.00 Median :12.52
## Mean :28.21 Mean : 9.128 Mean :30.68 Mean :12.33
## 3rd Qu.:32.81 3rd Qu.:11.050 3rd Qu.:33.41 3rd Qu.:14.74
## Max. :64.97 Max. :20.080 Max. :76.60 Max. :23.92
##
## TotalPctDiv PersPerFam PctFam2Par PctKids2Par
## Min. : 2.830 Min. :2.29 Min. :22.97 Min. :18.30
## 1st Qu.: 8.575 1st Qu.:2.99 1st Qu.:67.90 1st Qu.:63.99
## Median :10.900 Median :3.10 Median :75.03 Median :72.53

```

```

## Mean :10.813 Mean :3.13 Mean :74.06 Mean :71.23
## 3rd Qu.:12.985 3rd Qu.:3.22 3rd Qu.:81.90 3rd Qu.:80.39
## Max. :22.230 Max. :4.64 Max. :93.60 Max. :92.58
##
## PctYoungKids2Par PctTeen2Par PctWorkMomYoungKids PctWorkMom
## Min. : 8.70 Min. :20.20 Min. :24.42 Min. :41.95
## 1st Qu.: 74.78 1st Qu.:70.17 1st Qu.:55.43 1st Qu.:64.90
## Median : 83.99 Median :76.92 Median :60.71 Median :69.23
## Mean : 81.87 Mean :75.52 Mean :60.54 Mean :68.85
## 3rd Qu.: 91.67 3rd Qu.:82.77 3rd Qu.:65.98 3rd Qu.:73.50
## Max. :100.00 Max. :97.34 Max. :87.97 Max. :89.37
##
## NumKidsBornNeverMar PctKidsBornNeverMar NumImmig PctImmigRecent
## Min. : 0 Min. : 0.000 Min. : 20 Min. : 0.000
## 1st Qu.: 147 1st Qu.: 1.070 1st Qu.: 400 1st Qu.: 6.695
## Median : 352 Median : 2.040 Median : 1024 Median :12.260
## Mean : 2141 Mean : 3.115 Mean : 6277 Mean :13.526
## 3rd Qu.: 1032 3rd Qu.: 3.910 3rd Qu.: 3302 3rd Qu.:17.950
## Max. :527557 Max. :27.350 Max. :2082931 Max. :64.290
##
## PctImmigRec5 PctImmigRec8 PctImmigRec10 PctRecentImmig
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000
## 1st Qu.:11.26 1st Qu.:17.20 1st Qu.:22.73 1st Qu.: 0.170
## Median :19.08 Median :26.72 Median :34.79 Median : 0.500
## Mean :20.42 Mean :27.54 Mean :34.73 Mean : 1.099
## 3rd Qu.:27.45 3rd Qu.:36.49 3rd Qu.:46.19 3rd Qu.: 1.310
## Max. :76.16 Max. :80.81 Max. :88.00 Max. :13.710
##
## PctRecImmig5 PctRecImmig8 PctRecImmig10 PctSpeakEnglOnly
## Min. : 0.000 Min. : 0.000 Min. : 0.000 Min. : 6.15
## 1st Qu.: 0.280 1st Qu.: 0.390 1st Qu.: 0.520 1st Qu.:84.38
## Median : 0.750 Median : 1.040 Median : 1.310 Median :92.18
## Mean : 1.697 Mean : 2.308 Mean : 2.944 Mean :87.07
## 3rd Qu.: 2.015 3rd Qu.: 2.700 3rd Qu.: 3.455 3rd Qu.:95.45
## Max. :19.930 Max. :25.340 Max. :32.630 Max. :98.98
##
## PctNotSpeakEnglWell PctLargHouseFam PctLargHouseOccup PersPerOccupHous
## Min. : 0.000 Min. : 0.960 Min. : 0.440 Min. :1.580
## 1st Qu.: 0.510 1st Qu.: 3.390 1st Qu.: 2.370 1st Qu.:2.410
## Median : 0.920 Median : 4.280 Median : 3.050 Median :2.570
## Mean : 2.406 Mean : 5.387 Mean : 3.916 Mean :2.616
## 3rd Qu.: 2.270 3rd Qu.: 5.870 3rd Qu.: 4.210 3rd Qu.:2.770
## Max. :38.330 Max. :34.870 Max. :30.870 Max. :4.520
##
## PersPerOwnOccHous PersPerRentOccHous PctPersOwnOccup PctPersDenseHous
## Min. :1.61 Min. :1.550 Min. :13.93 Min. : 0.050
## 1st Qu.:2.55 1st Qu.:2.110 1st Qu.:57.28 1st Qu.: 1.290
## Median :2.71 Median :2.290 Median :65.91 Median : 2.340
## Mean :2.74 Mean :2.367 Mean :66.37 Mean : 4.132
## 3rd Qu.:2.90 3rd Qu.:2.530 3rd Qu.:76.58 3rd Qu.: 4.730
## Max. :4.48 Max. :4.730 Max. :97.24 Max. :59.490
##
## PctHousLess3BR MedNumBR HousVacant PctHousOccup
## Min. : 3.06 Min. :1.000 Min. : 36.0 Min. :37.47

```

##	1st Qu.:37.51	1st Qu.:2.000	1st Qu.: 304.5	1st Qu.:91.29
##	Median :46.39	Median :3.000	Median : 558.0	Median :94.21
##	Mean :45.41	Mean :2.641	Mean : 1748.4	Mean :92.93
##	3rd Qu.:53.52	3rd Qu.:3.000	3rd Qu.: 1228.0	3rd Qu.:96.02
##	Max. :95.34	Max. :4.000	Max. :172768.0	Max. :99.00
##				
##	PctHousOwnOcc	PctVacantBoarded	PctVacMore6Mos	MedYrHousBuilt
##	Min. :16.86	Min. : 0.000	Min. : 3.12	Min. :1939
##	1st Qu.:54.82	1st Qu.: 0.720	1st Qu.:24.48	1st Qu.:1956
##	Median :62.83	Median : 1.660	Median :34.10	Median :1964
##	Mean :63.37	Mean : 2.779	Mean :34.77	Mean :1963
##	3rd Qu.:72.64	3rd Qu.: 3.430	3rd Qu.:43.97	3rd Qu.:1971
##	Max. :96.49	Max. :39.890	Max. :82.13	Max. :1987
##				
##	PctHousNoPhone	PctWOFullPlumb	OwnOccLowQuart	OwnOccMedVal
##	Min. : 0.000	Min. :0.0000	Min. : 14999	Min. : 19500
##	1st Qu.: 0.905	1st Qu.:0.1600	1st Qu.: 41500	1st Qu.: 56200
##	Median : 2.850	Median :0.3200	Median : 65500	Median : 82800
##	Mean : 4.290	Mean :0.4253	Mean : 88696	Mean :113098
##	3rd Qu.: 6.805	3rd Qu.:0.5550	3rd Qu.:121500	3rd Qu.:150600
##	Max. :23.880	Max. :5.3300	Max. :500001	Max. :500001
##				
##	OwnOccHiQuart	OwnOccQrange	RentLowQ	RentMedian
##	Min. : 28200	Min. : 0	Min. : 99.0	Min. : 120.0
##	1st Qu.: 74300	1st Qu.: 32200	1st Qu.: 213.5	1st Qu.: 289.5
##	Median :106700	Median : 43400	Median : 307.0	Median : 397.0
##	Mean :145318	Mean : 56623	Mean : 330.0	Mean : 428.5
##	3rd Qu.:188000	3rd Qu.: 65450	3rd Qu.: 421.0	3rd Qu.: 544.0
##	Max. :500001	Max. :331000	Max. :1001.0	Max. :1001.0
##				
##	RentHighQ	RentQrange	MedRent	MedRentPctHousInc
##	Min. : 182.0	Min. : 0.0	Min. : 192.0	Min. :14.9
##	1st Qu.: 366.0	1st Qu.:139.0	1st Qu.: 364.0	1st Qu.:24.3
##	Median : 486.0	Median :171.0	Median : 467.0	Median :26.1
##	Mean : 527.3	Mean :197.3	Mean : 501.5	Mean :26.3
##	3rd Qu.: 659.5	3rd Qu.:232.5	3rd Qu.: 615.0	3rd Qu.:28.0
##	Max. :1001.0	Max. :803.0	Max. :1001.0	Max. :35.1
##				
##	MedOwnCostPctInc	MedOwnCostPctIncNoMtg	NumInShelters	NumStreet
##	Min. :14.00	Min. :10.10	Min. : 0.00	Min. : 0.00
##	1st Qu.:18.70	1st Qu.:12.00	1st Qu.: 0.00	1st Qu.: 0.00
##	Median :21.00	Median :12.80	Median : 0.00	Median : 0.00
##	Mean :20.99	Mean :13.01	Mean : 66.95	Mean : 17.82
##	3rd Qu.:23.10	3rd Qu.:13.70	3rd Qu.: 22.00	3rd Qu.: 1.00
##	Max. :32.70	Max. :23.40	Max. :23383.00	Max. :10447.00
##				
##	PctForeignBorn	PctBornSameState	PctSameHouse85	PctSameCity85
##	Min. : 0.18	Min. : 6.75	Min. :11.83	Min. :27.95
##	1st Qu.: 2.06	1st Qu.:50.11	1st Qu.:44.99	1st Qu.:72.06
##	Median : 4.31	Median :64.49	Median :52.17	Median :79.49
##	Mean : 7.34	Mean :61.54	Mean :51.54	Mean :77.41
##	3rd Qu.: 9.25	3rd Qu.:74.86	3rd Qu.:58.74	3rd Qu.:85.14
##	Max. :60.40	Max. :93.14	Max. :78.56	Max. :96.59
##				

```
## PctSameState85      LandArea      PopDens      PctUsePubTrans
## Min.   :32.83  Min.   : 0.90  Min.   : 10  Min.   : 0.000
## 1st Qu.:85.20  1st Qu.: 7.30  1st Qu.: 1182  1st Qu.: 0.360
## Median :90.03  Median : 13.70  Median : 2027  Median : 1.220
## Mean   :88.11  Mean   : 27.42  Mean   : 2784  Mean   : 3.041
## 3rd Qu.:93.01  3rd Qu.: 26.10  3rd Qu.: 3322  3rd Qu.: 3.365
## Max.   :99.90  Max.   :3569.80  Max.   :44230  Max.   :54.330
##
## PolicBudgPerPop      murders      murdPerPop      rapes
## Min.   : 15260  Min.   : 0.000  Min.   : 0.000  Min.   : 0.00
## 1st Qu.: 88094  1st Qu.: 0.000  1st Qu.: 0.000  1st Qu.: 2.00
## Median : 114582  Median : 1.000  Median : 2.170  Median : 7.00
## Mean   : 153578  Mean   : 7.765  Mean   : 5.859  Mean   : 28.05
## 3rd Qu.: 155656  3rd Qu.: 3.000  3rd Qu.: 8.365  3rd Qu.: 19.00
## Max.   :2422367  Max.   :1946.000  Max.   :91.090  Max.   :2818.00
## NA's   :1872                                NA's   :208
## rapesPerPop      robberies      robbbPerPop      assaults
## Min.   : 0.00  Min.   : 0.0  Min.   : 0.00  Min.   : 0.0
## 1st Qu.: 11.54  1st Qu.: 5.0  1st Qu.: 27.65  1st Qu.: 18.0
## Median : 26.92  Median : 19.0  Median : 74.80  Median : 56.0
## Mean   : 36.26  Mean   : 237.9  Mean   : 162.61  Mean   : 326.5
## 3rd Qu.: 51.47  3rd Qu.: 70.0  3rd Qu.: 187.16  3rd Qu.: 180.0
## Max.   :401.35  Max.   :86001.0  Max.   :2264.13  Max.   :62778.0
## NA's   :208  NA's   :1  NA's   :1  NA's   :13
## assaultPerPop      burglaries      burglPerPop      larcenies
## Min.   : 0.00  Min.   : 2.0  Min.   : 16.92  Min.   : 10
## 1st Qu.: 94.19  1st Qu.: 95.0  1st Qu.: 511.69  1st Qu.: 392
## Median : 226.53  Median : 205.0  Median : 822.72  Median : 747
## Mean   : 378.00  Mean   : 761.2  Mean   : 1033.43  Mean   : 2138
## 3rd Qu.: 504.39  3rd Qu.: 508.0  3rd Qu.: 1350.23  3rd Qu.: 1675
## Max.   :4932.50  Max.   :99207.0  Max.   :11881.02  Max.   :235132
## NA's   :13  NA's   :3  NA's   :3  NA's   :3
## larcPerPop      autoTheft      autoTheftPerPop      arsons
## Min.   : 77.86  Min.   : 1.0  Min.   : 6.55  Min.   : 0.00
## 1st Qu.: 2040.08  1st Qu.: 30.0  1st Qu.: 156.95  1st Qu.: 1.00
## Median : 3079.51  Median : 75.0  Median : 302.36  Median : 5.00
## Mean   : 3372.98  Mean   : 516.7  Mean   : 473.97  Mean   : 30.91
## 3rd Qu.: 4335.41  3rd Qu.: 232.5  3rd Qu.: 589.77  3rd Qu.: 16.00
## Max.   :25910.55  Max.   :112464.0  Max.   :4968.59  Max.   :5119.00
## NA's   :3  NA's   :3  NA's   :3  NA's   :91
## arsonsPerPop      ViolentCrimesPerPop      nonViolPerPop
## Min.   : 0.00  Min.   : 0.0  Min.   : 116.8
## 1st Qu.: 7.67  1st Qu.: 161.7  1st Qu.: 2918.1
## Median : 21.08  Median : 374.1  Median : 4425.4
## Mean   : 32.15  Mean   : 589.1  Mean   : 4908.2
## 3rd Qu.: 42.85  3rd Qu.: 794.4  3rd Qu.: 6229.3
## Max.   :436.37  Max.   :4877.1  Max.   :27119.8
## NA's   :91  NA's   :221  NA's   :97
```

```
str(crime2)
```

```
## 'data.frame': 2215 obs. of 123 variables:
## $ Êcommunityname : chr "BerkeleyHeightstownship" "Marpletownship" "Tigardcity" "Gloversville"
## $ state : chr "NJ" "PA" "OR" "NY" ...
```

```

## $ fold : int 1 1 1 1 1 1 1 1 1 1 ...
## $ population : int 11980 23123 29344 16656 11245 140494 28700 59459 74111 103590 ...
## $ householdsize : num 3.1 2.82 2.43 2.4 2.76 2.45 2.6 2.45 2.46 2.62 ...
## $ racepctblack : num 1.37 0.8 0.74 1.7 0.53 ...
## $ racePctWhite : num 91.8 95.6 94.3 97.3 89.2 ...
## $ racePctAsian : num 6.5 3.44 3.43 0.5 1.17 0.9 1.47 0.4 1.25 0.92 ...
## $ racePctHispanic : num 1.88 0.85 2.35 0.7 0.52 ...
## $ agePct12t21 : num 12.5 11 11.4 12.6 24.5 ...
## $ agePct12t29 : num 21.4 21.3 25.9 25.2 40.5 ...
## $ agePct16t24 : num 10.9 10.5 11 12.2 28.7 ...
## $ agePct65up : num 11.3 17.2 10.3 17.6 12.6 ...
## $ numbUrban : int 11980 23123 29344 0 0 140494 28700 59449 74115 103590 ...
## $ pctUrban : num 100 100 100 0 0 100 100 100 100 100 ...
## $ medIncome : int 75122 47917 35669 20580 17390 21577 42805 23221 25326 17852 ...
## $ pctWWage : num 89.2 79 82 68.2 69.3 ...
## $ pctWFarmSelf : num 1.55 1.11 1.15 0.24 0.55 1 0.39 0.67 2.93 0.86 ...
## $ pctWInvInc : num 70.2 64.1 55.7 39 42.8 ...
## $ pctWSocSec : num 23.6 35.5 22.2 39.5 32.2 ...
## $ pctWPubAsst : num 1.03 2.75 2.94 11.71 11.21 ...
## $ pctWRetire : num 18.4 22.9 14.6 18.3 14.4 ...
## $ medFamInc : int 79584 55323 42112 26501 24018 27705 50394 28901 34269 24058 ...
## $ perCapInc : int 29711 20148 16946 10810 8483 11878 18193 12161 13554 10195 ...
## $ whitePerCap : int 30233 20191 17103 10909 9009 12029 18276 12599 13727 12126 ...
## $ blackPerCap : int 13600 18137 16644 9984 887 7382 17342 9820 8852 5715 ...
## $ indianPerCap : int 5725 0 21606 4941 4425 10264 21482 6634 5344 11313 ...
## $ AsianPerCap : int 27101 20074 15528 3541 3352 10753 12639 8802 8011 5770 ...
## $ OtherPerCap : int 5115 5250 5954 2451 3000 7192 21852 7428 5332 7320 ...
## $ HispPerCap : int 22838 12222 8405 4391 1328 8104 22594 6187 5174 6984 ...
## $ NumUnderPov : int 227 885 1389 2831 2855 23223 1126 10320 9603 27767 ...
## $ PctPopUnderPov : num 1.96 3.98 4.75 17.23 29.99 ...
## $ PctLess9thGrade : num 5.81 5.61 2.8 11.05 12.15 ...
## $ PctNotHSGrad : num 9.9 13.72 9.09 33.68 23.06 ...
## $ PctBSorMore : num 48.2 29.9 30.1 10.8 25.3 ...
## $ PctUnemployed : num 2.7 2.43 4.01 9.86 9.08 5.72 4.85 8.19 4.18 8.39 ...
## $ PctEmploy : num 64.5 62 69.8 54.7 52.4 ...
## $ PctEmplManu : num 14.65 12.26 15.95 31.22 6.89 ...
## $ PctEmplProfServ : num 28.8 29.3 21.5 27.4 36.5 ...
## $ PctOccupManu : num 5.49 6.39 8.79 26.76 10.94 ...
## $ PctOccupMgmtProf : num 50.7 37.6 32.5 22.7 27.8 ...
## $ MalePctDivorce : num 3.67 4.23 10.1 10.98 7.51 ...
## $ MalePctNevMarr : num 26.4 28 25.8 28.1 50.7 ...
## $ FemalePctDiv : num 5.22 6.45 14.76 14.47 11.64 ...
## $ TotalPctDiv : num 4.47 5.42 12.55 12.91 9.73 ...
## $ PersPerFam : num 3.22 3.11 2.95 2.98 2.98 2.89 3.14 2.95 3 3.11 ...
## $ PctFam2Par : num 91.4 86.9 78.5 64 58.6 ...
## $ PctKids2Par : num 90.2 85.3 78.8 62.4 55.2 ...
## $ PctYoungKids2Par : num 95.8 96.8 92.4 65.4 66.5 ...
## $ PctTeen2Par : num 95.8 86.5 75.7 67.4 79.2 ...
## $ PctWorkMomYoungKids : num 44.6 51.1 66.1 59.6 61.2 ...
## $ PctWorkMom : num 58.9 62.4 74.2 70.3 68.9 ...
## $ NumKidsBornNeverMar : int 31 43 164 561 402 1511 263 2368 751 3537 ...
## $ PctKidsBornNeverMar : num 0.36 0.24 0.88 3.84 4.7 1.58 1.18 4.66 1.64 4.71 ...
## $ NumImmig : int 1277 1920 1468 339 196 2091 2637 517 1474 4793 ...
## $ PctImmigRecent : num 8.69 5.21 16.42 13.86 46.94 ...

```

```

## $ PctImmigRec5      : num 13 8.65 23.98 13.86 56.12 ...
## $ PctImmigRec8      : num 21 13.3 32.1 15.3 67.9 ...
## $ PctImmigRec10     : num 30.9 22.5 35.6 15.3 69.9 ...
## $ PctRecentImmig    : num 0.93 0.43 0.82 0.28 0.82 0.32 1.05 0.11 0.47 0.72 ...
## $ PctRecImmig5      : num 1.39 0.72 1.2 0.28 0.98 0.45 1.49 0.2 0.67 1.07 ...
## $ PctRecImmig8      : num 2.24 1.11 1.61 0.31 1.18 0.57 2.2 0.25 0.93 1.63 ...
## $ PctRecImmig10     : num 3.3 1.87 1.78 0.31 1.22 0.68 2.55 0.29 1.07 2.31 ...
## $ PctSpeakEnglOnly  : num 85.7 87.8 93.1 95 94.6 ...
## $ PctNotSpeakEnglWell : num 1.37 1.81 1.14 0.56 0.39 0.6 0.6 0.28 0.43 2.51 ...
## $ PctLargHouseFam    : num 4.81 4.25 2.97 3.93 5.23 3.08 5.08 3.85 2.59 6.7 ...
## $ PctLargHouseOccup  : num 4.17 3.34 2.05 2.56 3.11 1.92 3.46 2.55 1.54 4.1 ...
## $ PersPerOccupHous   : num 2.99 2.7 2.42 2.37 2.35 2.28 2.55 2.36 2.32 2.45 ...
## $ PersPerOwnOccHous  : num 3 2.83 2.69 2.51 2.55 2.37 2.89 2.42 2.77 2.47 ...
## $ PersPerRentOccHous : num 2.84 1.96 2.06 2.2 2.12 2.16 2.09 2.27 1.91 2.44 ...
## $ PctPersOwnOccup    : num 91.5 89 64.2 58.2 58.1 ...
## $ PctPersDenseHous   : num 0.39 1.01 2.03 1.21 2.94 2.11 1.47 1.9 1.67 6.14 ...
## $ PctHousLess3BR     : num 11.1 23.6 47.5 45.7 55.6 ...
## $ MedNumBR           : int 3 3 3 3 2 2 3 2 2 2 ...
## $ HousVacant         : int 64 240 544 669 333 5119 566 2051 1562 5606 ...
## $ PctHousOccup       : num 98.4 97.2 95.7 91.2 92.5 ...
## $ PctHousOwnOcc      : num 91 84.9 57.8 54.9 53.6 ...
## $ PctVacantBoarded   : num 3.12 0 0.92 2.54 3.9 2.09 1.41 6.39 0.45 5.64 ...
## $ PctVacMore6Mos     : num 37.5 18.33 7.54 57.85 42.64 ...
## $ MedYrHousBuilt     : int 1959 1958 1976 1939 1958 1966 1956 1954 1971 1960 ...
## $ PctHousNoPhone     : num 0 0.31 1.55 7 7.45 ...
## $ PctWOFullPlumb     : num 0.28 0.14 0.12 0.87 0.82 0.31 0.28 0.49 0.19 0.33 ...
## $ OwnOccLowQuart     : int 215900 136300 74700 36400 30600 37700 155100 26300 54500 28600 ...
## $ OwnOccMedVal       : int 262600 164200 90400 49600 43200 53900 179000 37000 70300 43100 ...
## $ OwnOccHiQuart      : int 326900 199900 112000 66500 59500 73100 215500 52400 93700 67400 ...
## $ OwnOccQrange       : int 111000 63600 37300 30100 28900 35400 60400 26100 39200 38800 ...
## $ RentLowQ           : int 685 467 370 195 202 215 463 186 241 192 ...
## $ RentMedian         : int 1001 560 428 250 283 280 669 253 321 281 ...
## $ RentHighQ          : int 1001 672 520 309 362 349 824 325 387 369 ...
## $ RentQrange         : int 316 205 150 114 160 134 361 139 146 177 ...
## $ MedRent            : int 1001 627 484 333 332 340 736 338 355 353 ...
## $ MedRentPctHousInc  : num 23.8 27.6 24.1 28.7 32.2 26.4 24.4 26.3 25.2 29.6 ...
## $ MedOwnCostPctInc   : num 21.1 20.7 21.7 20.6 23.2 17.3 20.8 15.1 20.7 19.4 ...
## $ MedOwnCostPctIncNoMtg : num 14 12.5 11.6 14.5 12.9 11.7 12.5 12.2 12.8 13 ...
## $ NumInShelters      : int 11 0 16 0 2 327 0 21 125 43 ...
## $ NumStreet          : int 0 0 0 0 0 4 0 0 15 4 ...
## $ PctForeignBorn     : num 10.66 8.3 5 2.04 1.74 ...
## $ PctBornSameState   : num 53.7 77.2 44.8 88.7 73.8 ...
## $ PctSameHouse85     : num 65.3 71.3 36.6 56.7 42.2 ...
## [list output truncated]

```

```

# The result of 2215 records and 123 variables
# match with the missing data treatment steps. No
# record was deleted, but 24 variables were
# removed out of the 147 variables.

```

```

# Outlier Analysis

```

```

# Assuming that RentMedian is the Y variable in

```



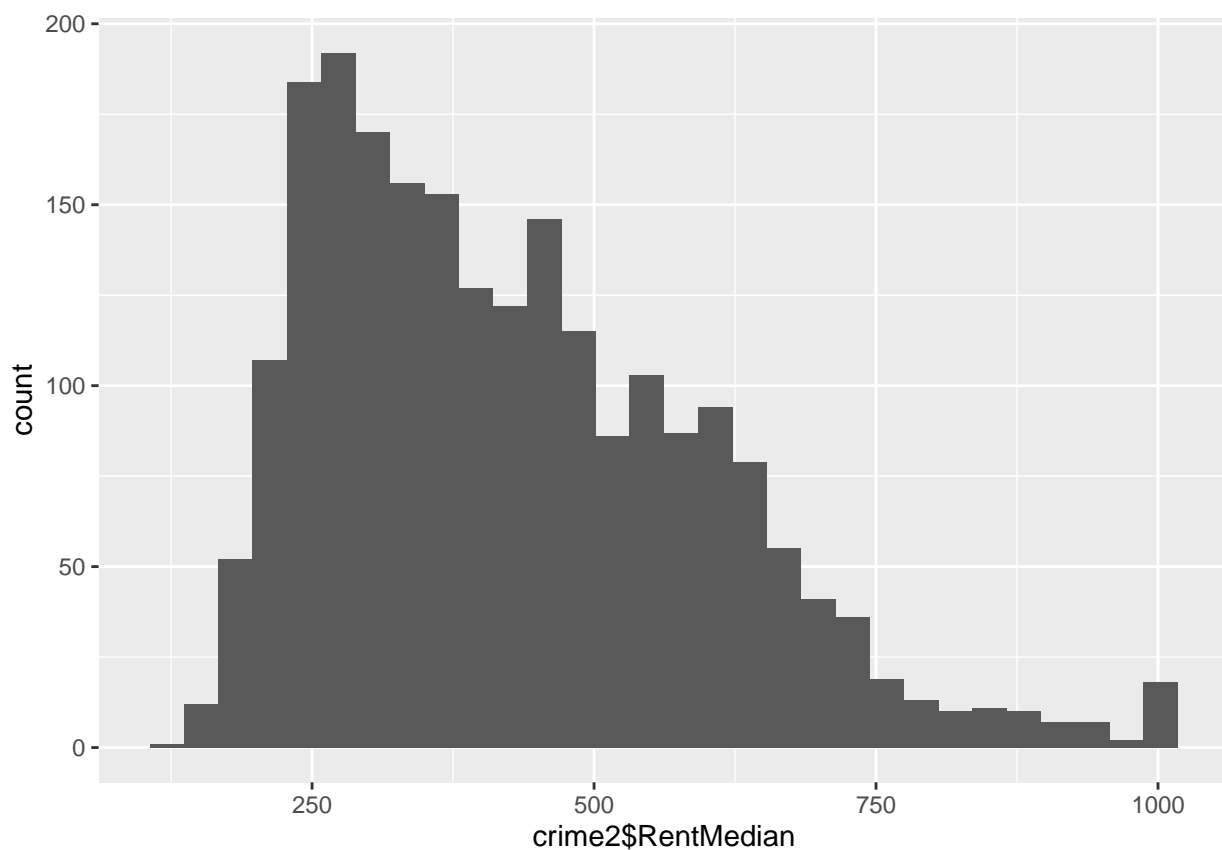
```
# this analysis, let us look at the number of  
# outliers for this dependent variable. Firstly  
# let us check the histogram and box plot for a  
# general idea.
```

```
library(ggplot2)
```

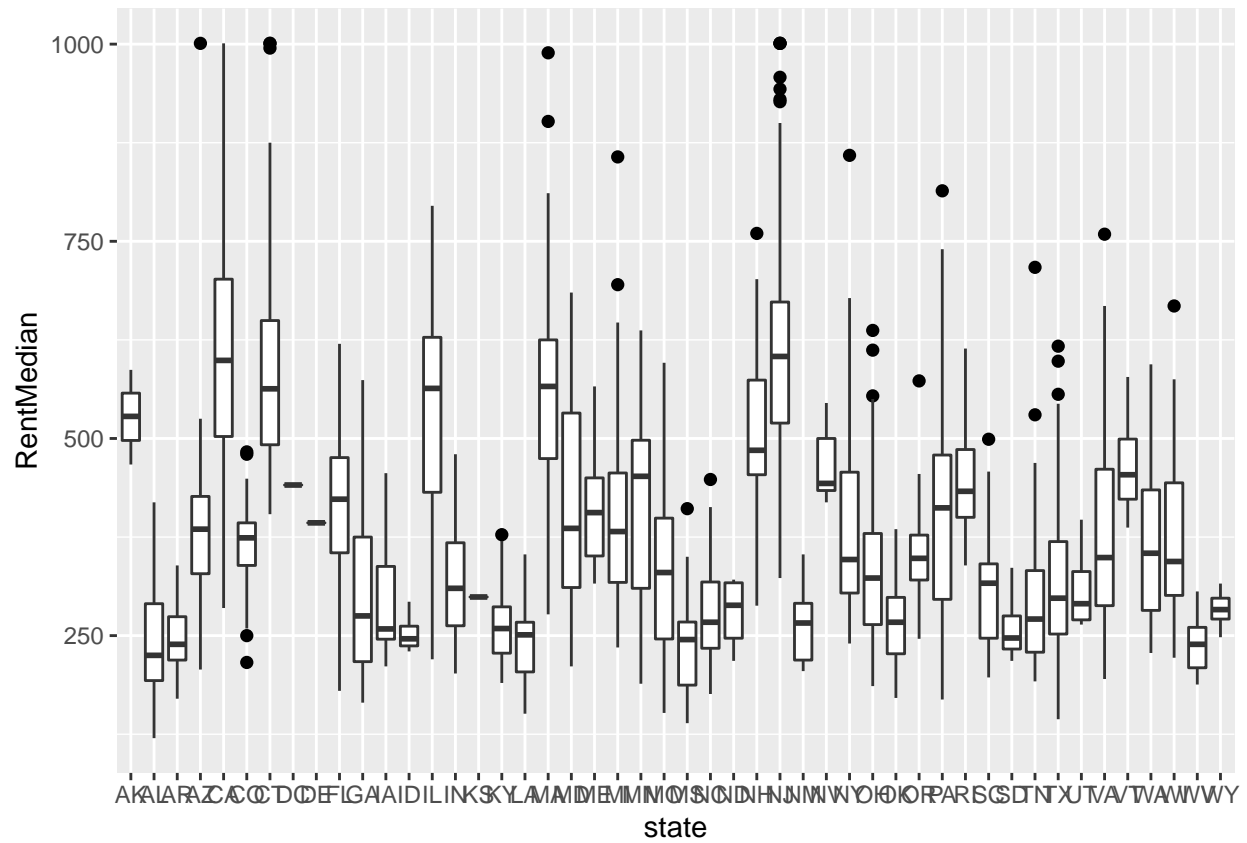
```
ggplot(data = crime2, aes(crime2$RentMedian)) + geom_histogram()
```

```
## Warning: Use of 'crime2$RentMedian' is discouraged. Use 'RentMedian' instead.
```

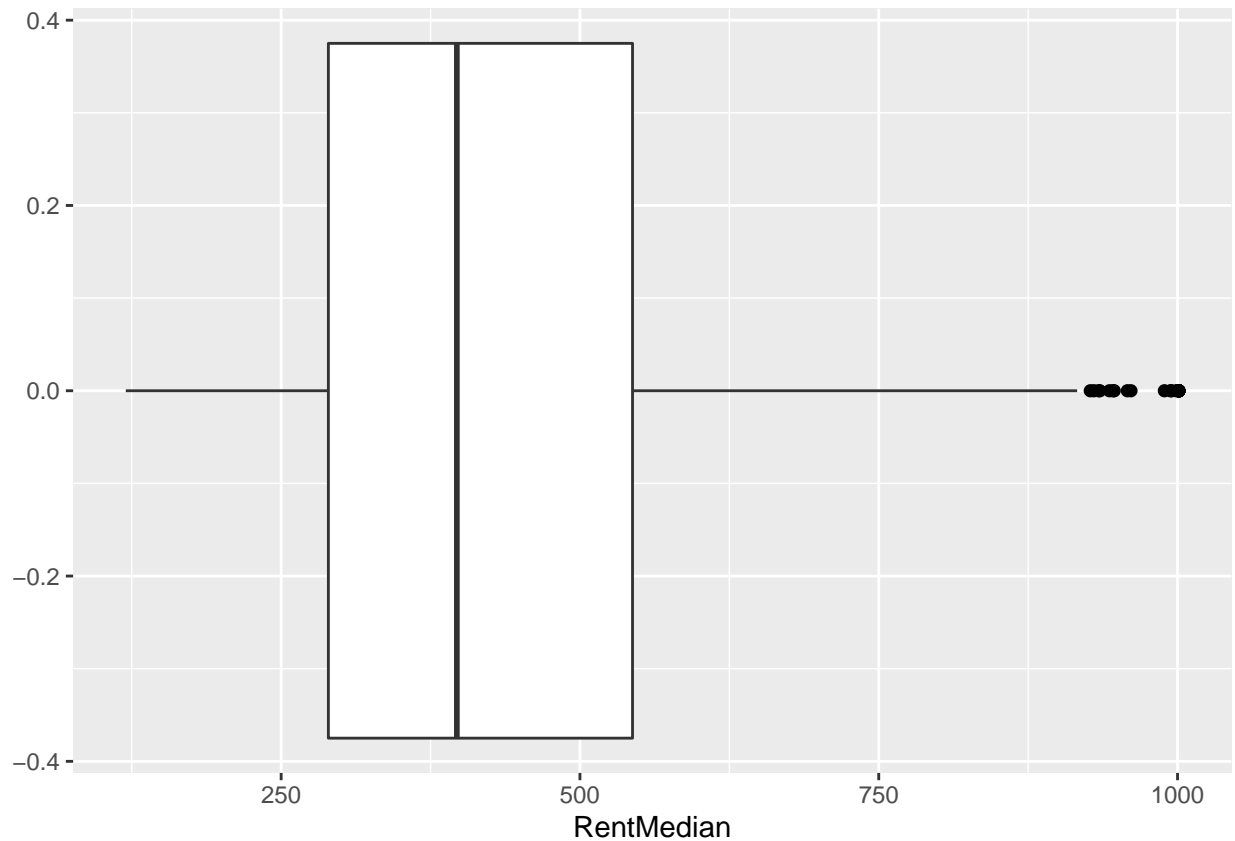
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
ggplot(data = crime2, aes(x = state, y = RentMedian)) +  
  geom_boxplot(outlier.colour = "black", outlier.shape = 16,  
    outlier.size = 2, notch = FALSE)
```



```
ggplot(data = crime2, aes(RentMedian)) + geom_boxplot(outlier.colour = "black",
  outlier.shape = 16, outlier.size = 2, notch = FALSE)
```



*# Histogram and Boxplot show outliers on the
 # right tail of the data for the variable
 # RentMedian. If we take a closer look, for each
 # state the outliers if any usually appear on the
 # higher end of the spectrum. It indicates that
 # there are luxury properties or properties of
 # unique features/locations etc. that make them
 # stick out in pricing far above the mean. To
 # quantify this, let us examine the z scores of
 # the records and check how many outliers there
 # are.*

```
z <- (crime2$RentMedian - mean(crime2$RentMedian))/sd(crime2$RentMedian)
```

```
number_of_outliers_for_Y <- sum(abs(z) > 3)
```

```
ratio_of_outliers_for_Y <- number_of_outliers_for_Y/2215
```

```
number_of_outliers_for_Y
```

```
## [1] 23
```

```
ratio_of_outliers_for_Y
```

```
## [1] 0.01038375
```

```
# The results show that 23 data points out of  
# 2215 records for RentMedian are outliers  
# because they are more than 3 standard  
# deviations away from the mean. The percentage  
# of outliers for the defined Y is around 1.04%.
```