

# 고뇌하는 4조



범죄는  
어디서 얼마나  
일어나는가  
In America



E

D

A



역할분담

곽현지

신혜연



“클러스터링”



주일 찬

장은수



“데이터 클리닝  
& PPT 제작”

하민진

정권상



“Feature Engineering”

# 목 차



1

A GLANCE

DATA CLEANSING

3

CLUSTERING

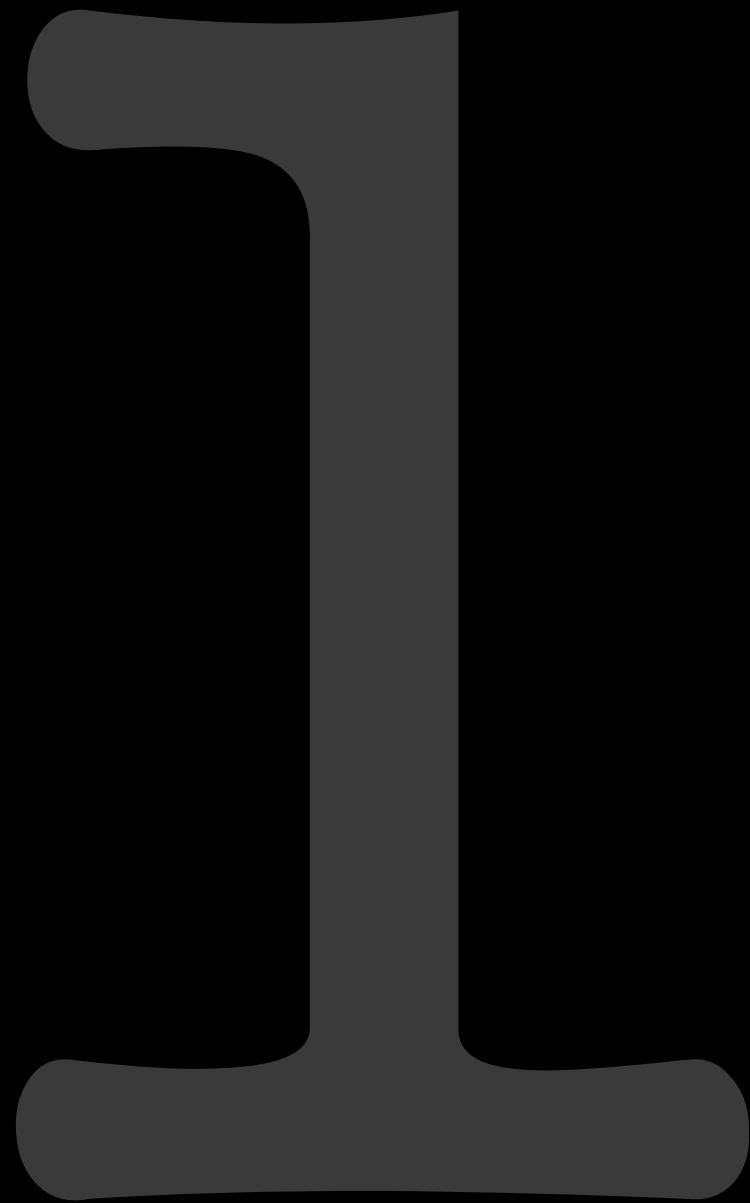
TRANSFORMATION

2

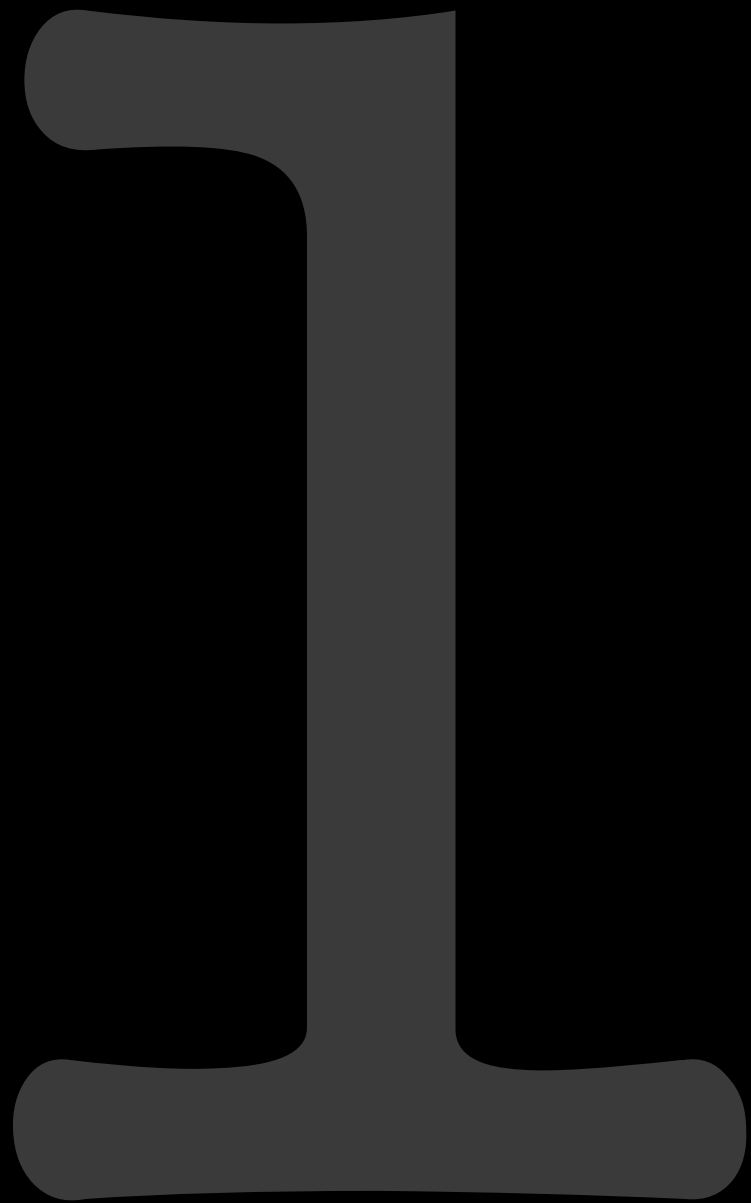
4

A large, dark gray, stylized letter 'I' with rounded ends, positioned on the left side of the image.

탐색전 : 데이터 둘러보기



“쓰-”



3	Marpletown	PA	45	47616	1	23123	2.82	0.8	95.57	3.44	0.85	11.01	21.3	10.48	17.18	23123	100	47917	78.9
4	Tigard	city OR	?	?	1	29344	2.43	0.74	94.33	3.43	2.35	11.36	25.88	11.01	10.28	29344	100	35669	8
5	Gloversville	city NY	35	29443	1	16656	2.4	1.7	97.35	0.5	0.7	12.55	25.2	12.19	17.57	0	0	20580	68.1
6	Bemidji	city MN	7	5068	1	11245	2.76	0.53	89.16	1.17	0.52	24.46	40.53	28.69	12.65	0	0	17390	69.3
7	Springfield	city MO	?	?	1	140494	2.45	2.51	95.65	0.9	0.95	18.09	32.89	20.04	13.26	140494	100	21577	75.7
8	Norwood	town MA	21	50250	1	28700	2.6	1.6	96.57	1.47	1.1	11.17	27.41	12.76	14.42	28700	100	42805	79.4
9	Anderson	city IN	?	?	1	59459	2.45	14.2	84.87	0.4	0.63	15.31	27.93	14.78	14.6	59449	100	23221	71
10	Fargo	city ND	17	25700	1	74111	2.46	0.35	97.11	1.25	0.73	16.64	35.16	20.33	8.58	74115	100	25326	83.6
11	Waco	city TX	?	?	1	103590	2.62	23.14	67.6	0.92	16.35	19.88	34.55	21.62	13.12	103590	100	17852	74
12	Sherman	city TX	?	?	1	31601	2.54	12.63	83.22	0.77	4.39	15.73	28.57	15.16	14.26	31596	100	24763	73.9
13	San Pablo	city CA	?	?	1	25158	2.89	21.34	49.42	17.21	26.78	13.65	28.82	13.23	9.44	25158	100	25479	73.4
14	Bowling Green	city KY	?	?	1	40641	2.54	12.18	86.39	1.12	0.68	21.51	36.83	23.96	11.5	0	0	20043	75.2
15	Pine Bluff	city AR	?	?	1	57140	2.74	53.52	45.65	0.49	0.43	16.51	28.17	14.68	13.38	57140	100	19143	69.3
16	New Ulm	city MN	15	46042	1	13132	2.53	0.06	99.21	0.47	0.59	14	25.03	12.83	14.57	0	0	25797	73.9
17	Maplewood	city MN	123	40382	1	30954	2.69	2.52	94.39	2.03	1.55	12.07	25.43	11.42	10.45	30954	100	37856	82.2
18	Enfield	town CT	3	25990	1	45532	2.85	2.65	95.72	1.04	2.28	11.86	27.51	12.36	9.76	43944	96.51	44635	84.9
19	Glendale	city CA	?	?	1	180038	2.62	1.3	74.02	14.14	20.96	12.04	26.68	12.37	11.54	180038	100	34372	76.1
20	Worthington	city OH	?	?	1	14869	2.67	2.28	94.74	2.67	0.74	13.71	20.33	9.48	12.38	14882	100	49851	81.8
21	Arlington	city TX	?	?	1	261721	2.6	8.41	82.64	3.92	8.91	14.18	32.78	15.14	4.58	261763	100	35048	90.2
22	Plymouth	city MN	53	51730	1	50889	2.77	1.61	95.66	2.04	1.02	13.13	26.94	12.19	4.77	50889	100	51314	92.2
23	New York	city NY	?	?	1	7322564	2.6	28.71	52.26	7	24.36	13.06	27.46	13.09	11.62	7322564	100	29823	73.5
24	Marina	city CA	?	?	1	26436	3.34	18.97	53.6	20.84	10.73	16.16	37.22	19.09	4.13	26436	100	29043	90.3
25	Lebanon	city NH	9	41300	1	12183	2.36	0.41	97.55	1.55	0.91	11.29	26.87	12.2	11.28	0	0	32221	84.6
26	Rockledge	city FL	?	?	1	16023	2.63	13.79	83.94	1.42	2.4	11.91	23.09	10.33	13.65	16023	100	34934	80.4
27	Rogers	city AR	?	?	1	24692	2.54	0.06	97.72	0.77	1.86	12.84	25.81	11.93	14.62	0	0	26198	73.9
28	Bellaire	city TX	?	?	1	13842	2.35	0.41	94.65	1.98	7.95	8.23	18.3	6.98	13.57	13842	100	45892	80.4
29	El Cajon	city CA	?	?	1	88693	2.7	2.92	87.36	2.82	13.97	13.77	30.92	15.15	9.65	88693	100	28108	78.1
30	Moses Lake	city WA	?	?	1	11235	2.6	1.89	82.45	1.82	18.16	14.59	25.62	12.02	13.36	0	0	23258	72.4
31	West Memphis	city AR	?	?	1	28259	2.86	42.15	56.94	0.52	0.5	16.26	28.79	13.83	9.53	28259	100	22052	77.2
32	Eunice	city LA	?	?	1	11162	2.8	26.96	72.26	0.4	0.78	15.43	26.03	12.04	12.78	0	0	14874	62.8
33	Laredo	city TX	?	?	1	122899	3.84	0.12	70.83	0.38	93.87	19.93	33.12	16.79	7.32	122899	100	18395	79.1
34	Amsterdam	city NY	57	2066	1	20714	2.36	1.46	93.15	0.56	11.61	11.35	23.13	11.19	21.15	0	0	22166	6
35	Gorham	town ME	5	28240	1	11856	3.03	0.37	98.84	0.4	0.45	20.11	31.45	19.01	8.63	3929	33.1	36618	83.2
36	Rock Springs	city WY	?	?	1	19050	2.67	1.17	94.31	0.96	7.46	15.43	26.48	11.73	8.64	0	0	34372	83.5
37	Oakdale	city CA	2	2	1	11961	2.71	0.24	88.57	1.11	17.04	13.37	26.84	13.03	12.08	0	0	27220	72

3	Marpletowns	PA		45	47616	1	23123	2.82	0.8	95.57	3.44	0.85	11.01	21.3	10.48	17.18	23123	100	47917	78.9
4	Tigard	city OR	?	?		1	29344	2.43	0.74	94.33	3.43	2.35	11.36	25.88	11.01	10.28	29344	100	35669	8
5	Gloversville	city NY		35	29443	1	16656	2.4	1.7	97.35	0.5	0.7	12.55	25.2	12.19	17.57	0	0	20580	68.1
6	Bemidji	city MN		7	5068	1	11245	2.76	0.53	89.16	1.17	0.52	24.46	40.53	28.69	12.65	0	0	17390	69.3
7	Springfield	city MO	?	?		1	140494	2.45	2.51	95.65	0.9	0.95	18.09	32.89	20.04	13.26	140494	100	21577	75.7
8	Norwood	town MA		21	50250	1	28700	2.6	1.6	96.57	1.47	1.1	11.17	27.41	12.76	14.42	28700	100	42805	79
9	Anderson	city IN	?	?		1	59459	2.45	14.2	84.87	0.4	0.63	15.31	27.93	14.78	14.6	59449	100	23221	71
10	Fargo	city ND		17	25700	1	74111	2.46	0.35	97.11	1.25	0.73	16.64	35.16	20.33	8.58	74115	100	25326	83.6
11	Waco	city TX	?	?		1	103590	2.62	23.14	67.6	0.92	16.35	19.88	34.55	21.62	13.12	103590	100	17852	74
12	Sherman	city TX	?	?		1	31601	2.54	12.63	83.22	0.77	4.39	15.73	28.57	15.16	14.26	31596	100	24763	73.9
13	San Pablo	city CA	?	?		1	25158	2.89	21.34	49.42	17.21	26.78	13.65	28.82	13.23	9.44	25158	100	25479	73.4
14	Bowling Green	KY	?	?		1	40641	2.54	12.18	86.39	1.12	0.68	21.51	36.83	23.96	11.5	0	0	20043	75.2
15	Pine Bluff	city AR	?	?		1	57140	2.74	53.52	45.65	0.49	0.43	16.51	28.17	14.68	13.38	57140	100	19143	69.3
16	New Ulm	city MN		15	46042	1	13132	2.53	0.06	99.21	0.47	0.59	14	25.03	12.83	14.57	0	0	25797	73.9
17	Maplewood	city MN		123	40382	1	30954	2.69	2.52	94.39	2.03	1.55	12.07	25.43	11.42	10.45	30954	100	37856	82.2
18	Enfield	town CT		3	25990	1	45532	2.85	2.65	95.72	1.04	2.28	11.86	27.51	12.36	9.76	43944	96.51	44635	84.9
19	Glendale	city CA	?	?		1	18038	2.62	1.3	74.02	1.1	20.96	12.04	26.65	12.37	11.24	180038	100	34372	76.1
20	Worthington	city OH	?	?		1	14869	2.67	2.28	94.74	0.57	0.1	13.11	29.03	12.38	14.882	14882	100	49851	81.8
21	Arlington	city TX	?	?		1	261721	2.6	8.41	82.64	3.92	8.91	14.18	32.78	15.14	4.58	261763	100	35048	90.2
22	Plymouth	city MN		53	51730	1	50889	2.77	1.61	95.66	2.04	1.02	13.13	26.94	12.19	4.77	50889	100	51314	92.2
23	New York	city NY	?	?		1	7322564	2.6	28.71	52.26	7	24.36	13.06	27.46	13.09	11.62	7322564	100	29823	73.5
24	Marina	city CA	?	?		1	26436	3.34	18.97	53.6	20.84	10.73	16.16	37.22	19.09	4.13	26436	100	29043	90.3
25	Lebanon	city NH		9	41300	1	12183	2.36	0.41	97.55	1.55	0.91	11.29	26.87	12.2	11.28	0	0	32221	84.6
26	Rockledge	city FL	?	?		1	16023	2.63	13.79	83.94	1.42	2.4	11.91	23.09	10.33	13.65	16023	100	34934	80.1
27	Rogers	city AR	?	?		1	24692	2.54	0.06	97.72	0.77	1.86	12.84	25.81	11.93	14.62	0	0	26198	73.9
28	Bellaire	city TX	?	?		1	13842	2.35	0.41	94.65	1.98	7.95	8.23	18.3	6.98	13.57	13842	100	45892	80.4
29	El Cajon	city CA	?	?		1	88693	2.7	2.92	87.36	2.82	13.97	13.77	30.92	15.15	9.65	88693	100	28108	78.1
30	Moses Lake	city WA	?	?		1	11235	2.6	1.89	82.45	1.82	18.16	14.59	25.62	12.02	13.36	0	0	23258	72.4
31	West Memphis	AR	?	?		1	28259	2.86	42.15	56.94	0.52	0.5	16.26	28.79	13.83	9.53	28259	100	22052	77.2
32	Eunice	city LA	?	?		1	11162	2.8	26.96	72.26	0.4	0.78	15.43	26.03	12.04	12.78	0	0	14874	62.8
33	Laredo	city TX	?	?		1	122899	3.84	0.12	70.83	0.38	93.87	19.93	33.12	16.79	7.32	122899	100	18395	79.1
34	Amsterdam	city NY		57	2066	1	20714	2.36	1.46	93.15	0.56	11.61	11.35	23.13	11.19	21.15	0	0	22166	6
35	Gorham	town ME		5	28240	1	11856	3.03	0.37	98.84	0.4	0.45	20.11	31.45	19.01	8.63	3929	33.1	36618	83.2
36	Rock Springs	city WY	?	?		1	19050	2.67	1.17	94.31	0.96	7.46	15.43	26.48	11.73	8.64	0	0	34372	83.5
37	Oakdale	city CA	?	?		1	11061	2.71	0.24	88.57	1.11	17.04	12.27	26.84	12.02	12.08	0	0	27220	72

“2215 obs of 147 variables”

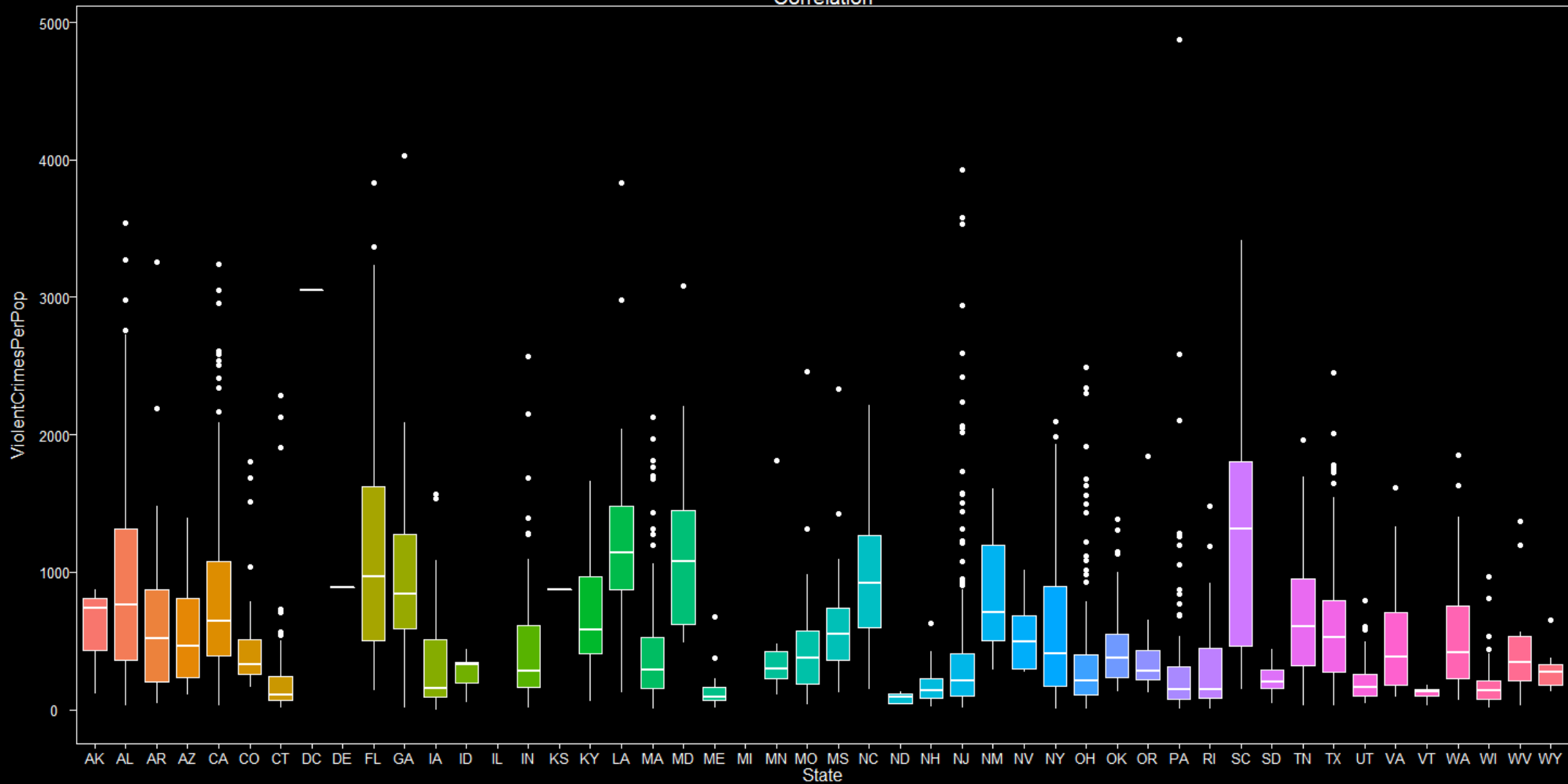


3	Marpletownsl	PA		45	47616	1	23123	2.82	0.8	95.57	3.44	0.85	11.01	21.3	10.48	17.18	23123	100	47917	78.9
4	Tigardcity	OR	?	?		1	29344	2.43	0.74	94.33	3.43	2.35	11.36	25.88	11.01	10.28	29344	100	35669	8
5	Gloversvillecit	NY		35	29443	1	16656	2.4	1.7	97.35	0.5	0.7	12.55	25.2	12.19	17.57	0	0	20580	68.1
6	Bemidjicity	MN		7	5068	1	11245	2.76	0.53	89.16	1.17	0.52	24.46	40.53	28.69	12.65	0	0	17390	69.3
7	Springfieldcit	MO	?	?		1	140494	2.45	2.51	95.65	0.9	0.95	18.09	32.89	20.04	13.26	140494	100	21577	75.7
8	Norwoodtown	MA		21	50250	1	28700	2.6	1.6	96.57	1.47	1.1	11.17	27.41	12.76	14.42	28700	100	42805	79
9	Andersoncity	IN	?	?		1	59459	2.45	14.2	84.87	0.4	0.63	15.31	27.93	14.78	14.6	59449	100	23221	71
10	Fargocity	ND		17	25700	1	74111	2.46	0.35	97.11	1.25	0.73	16.64	35.16	20.33	8.58	74115	100	25326	83.6
11	Wacocity	TX	?	?		1	103590	2.62	23.14	67.6	0.92	16.35	19.88	34.55	21.62	13.12	103590	100	17852	74
12	Shermancity	TX	?	?		1	31601	2.54	12.63	83.22	0.77	4.39	15.73	28.57	15.16	14.26	31596	100	24763	73.9
13	SanPablocity	CA	?	?		1	25158	2.89	21.34	49.42	17.21	26.78	13.65	28.82	13.23	9.44	25158	100	25479	73.4
14	BowlingGreer	KY	?	?		1	40641	2.54	12.18	86.39	1.12	0.68	21.51	36.83	23.96	11.5	0	0	20043	75.2
15	PineBluffcity	AR	?	?		1	57140	2.74	53.52	45.65	0.49	0.43	16.51	28.17	14.68	13.38	57140	100	19143	69.3
16	NewUlmcity	MN		15	46042	1	13132	2.53	0.06	99.21	0.47	0.59	14	25.03	12.83	14.57	0	0	25797	73.9
17	Maplewoodci	MN		123	40382	1	30954	2.69	2.52	94.39	2.03	1.55	12.07	25.43	11.42	10.45	30954	100	37856	82.2
18	Enfieldtown	CT		3	25990	1	45532	2.85	2.65	95.72	1.04	2.28	11.86	27.51	12.36	9.76	43944	96.51	44635	84.9
19	Glendalecity	CA	?	?		1	180038	2.62	1.3	75.02	1.7	20.96	12.4	5.6	22.7	11.54	180038	100	34372	76.1
20	Worthingtonc	OH	?	?		1	14166	2.28	1.74	86.74	0.74	0.74	22.3	29.48	12.38	14882	100	49851	81.8	
21	Arlingtoncity	TX	?	?		1	261721	2.6	8.41	82.64	3.92	8.91	14.18	32.78	15.14	4.58	261763	100	35048	90.2
22	Plymouthcity	MN		53	51730	1	50889	2.77	1.61	95.66	2.04	1.02	13.13	26.94	12.19	4.77	50889	100	51314	92.2
23	NewYorkcity	NY	?	?		1	7322564	2.6	28.71	52.26	7	24.36	13.06	27.46	13.09	11.62	7322564	100	29823	73.5
24	Marinacity	CA	?	?		1	26436	3.34	18.97	53.6	20.84	10.73	16.16	37.22	19.09	4.13	26436	100	29043	90.3
25	Lebanoncity	NH		9	41300	1	12183	2.36	0.41	97.55	1.55	0.91	11.29	26.87	12.2	11.28	0	0	32221	84.6
26	Rockledgecity	FL	?	?		1	16023	2.63	13.79	83.94	1.42	2.4	11.91	23.09	10.33	13.65	16023	100	34934	80
27	Rogerscity	AR	?	?		1	24692	2.54	0.06	97.72	0.77	1.86	12.84	25.81	11.93	14.62	0	0	26198	73.9
28	Bellairecity	TX	?	?		1	13842	2.35	0.41	94.65	1.98	7.95	8.23	18.3	6.98	13.57	13842	100	45892	80.4
29	ElCajoncity	CA	?	?		1	88693	2.7	2.92	87.36	2.82	13.97	13.77	30.92	15.15	9.65	88693	100	28108	78.1
30	MosesLakecit	WA	?	?		1	11235	2.6	1.89	82.45	1.82	18.16	14.59	25.62	12.02	13.36	0	0	23258	72.4
31	WestMemphi	AR	?	?		1	28259	2.86	42.15	56.94	0.52	0.5	16.26	28.79	13.83	9.53	28259	100	22052	77.2
32	Eunicecity	LA	?	?		1	11162	2.8	26.96	72.26	0.4	0.78	15.43	26.03	12.04	12.78	0	0	14874	62.8
33	Laredocity	TX	?	?		1	122899	3.84	0.12	70.83	0.38	93.87	19.93	33.12	16.79	7.32	122899	100	18395	79.1
34	Amsterdamci	NY		57	2066	1	20714	2.36	1.46	93.15	0.56	11.61	11.35	23.13	11.19	21.15	0	0	22166	6
35	Gorhamtown	ME		5	28240	1	11856	3.03	0.37	98.84	0.4	0.45	20.11	31.45	19.01	8.63	3929	33.1	36618	83.2
36	RockSpringsc	WY	?	?		1	19050	2.67	1.17	94.31	0.96	7.46	15.43	26.48	11.73	8.64	0	0	34372	83.5
37	Oakdalecity	CA	?	?		1	11061	2.71	0.24	88.57	1.11	17.04	12.27	26.84	12.02	12.08	0	0	27220	72

“number of NA : 44592”



Correlation



communityname	state	countyCode	communityCo	fold
BerkeleyHeightstown	NJ	39	5320	1
Marpletownship	PA	45	47616	1
Tigardcity	OR	?	?	1
Gloversvillecity	NY	35	29443	1
Bemidjicity	MN	7	5068	1
Springfieldcity	MO	?	?	1
Norwoodtown	MA	21	50250	1
Andersoncity	IN	?	?	1
Fargocity	ND	17	25700	1
Wacocity	TX	?	?	1
Shermancity	TX	?	?	1
SanPablocity	CA	?	?	1
BowlingGreencity	KY	?	?	1
PineBluffcity	AR	?	?	1
NewUlmcity	MN	15	46042	1
Maplewoodcity	MN	123	40382	1
Enfieldtown	CT	3	25990	1
Glendalecity	CA	?	?	1
Worthingtoncity	OH	?	?	1
Arlingtoncity	TX	?	?	1
Plymouthcity	MN	53	51730	1

## 팩터화

state	countyCode	communityCo
NJ	39	5320
PA	45	47616
OR	?	?
NY	35	29443
MN	7	5068
MO	?	?
MA	21	50250
IN	?	?
ND	17	25700
TX	?	?
TX	?	?
CA	?	?
KY	?	?
AR	?	?
MN	15	46042
MN	123	40382
CT	3	25990
CA	?	?
OH	?	?
TX	?	?
MN	53	51730

```
sum(train == '?')  
train[train=='?'] <- NA
```

## Numeric화

```
for(i in 3:dim(train)[2]){  
  if(is.character(train[[i]])){  
    train[[i]] <- as.numeric(train[[i]])  
  }  
}
```

1

“0”

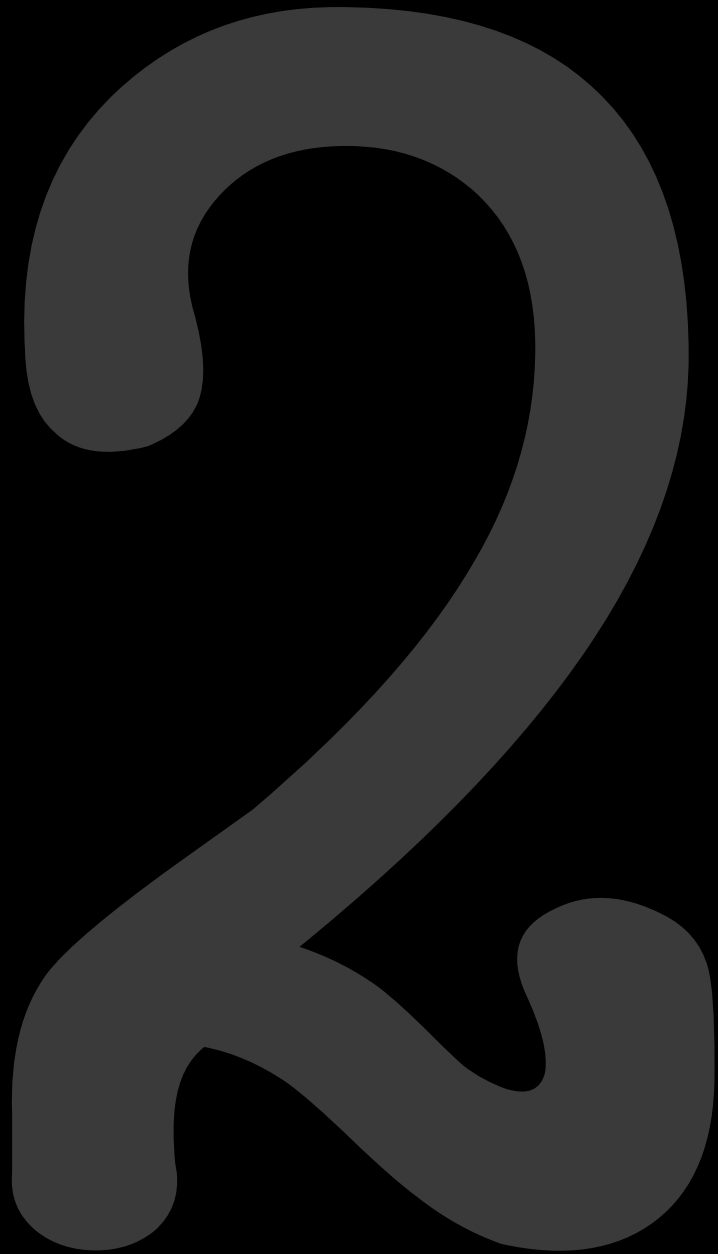
2

“0, 째이나 찢이나”

AB
blackPerCap
13600
18137
16644
9984
887
7382
10412
6016
5688
0
13140
8297
17693

( 예 시 )

EB
rapes
0
1
6
10
?
77
4
34
35
141
29
21
36
59



CLEANSING



2

REMOVE

1

# 코드 변수 삭제

( 'countyCode' , 'communityCode' )

A	B	C	D
community	state	countyCoc	community
BerkeleyH	NJ	39	5320
Marpletow	PA	45	47616
Tigardcity	OR	?	?
Gloversvill	NY	35	29443
Bemidjcity	MN	7	5068
Springfiel	MO	?	?
Norwoodt	MA	21	50250
Andersonc	IN	?	?
Fargocity	ND	17	25700
Wacocity	TX	?	?
Shermanci	TX	?	?
SanPabloc	CA	?	?
BowlingGr	KY	?	?
PineBluffci	AR	?	?
NewUlmci	MN	15	46042
Maplewoc	MN	123	40382
Enfieldtow	CT	3	25990
Glendaleci	CA	?	?
Worthingt	OH	?	?
Arlingtonc	TX	?	?
Plymouthc	MN	53	51730



A	B	C	D
community	state	countyCoc	community
BerkeleyH	NJ	39	5320
Marpletow	PA	45	47616
Tigardcity	OR	?	?
Gloversvill	NY	35	29443
Bemidjcity	MN	7	5068
Springfiel	MO	?	?
Norwoodt	MA	21	50250
Andersonc	IN	?	?
Fargocity	ND	17	25700
Wacocity	TX	?	?
Shermanci	TX	?	?
SanPabloc	CA	?	?
BowlingGr	KY	?	?
PineBluffci	AR	?	?
NewUlmci	MN	15	46042
Maplewoc	MN	123	40382
Enfieldtow	CT	3	25990
Glendaleci	CA	?	?
Worthingt	OH	?	?
Arlingtonc	TX	?	?
Plymouthc	MN	53	51730
NewYorkci	NY	?	?

# 정보 중복 변수 삭제

예시)

ED	EE
robberies	robbbPerP
1	8.2
5	21.26
56	154.95
10	57.86
4	32.04

‘robberies’ : 강도 수

‘robbbPerPop’ : 강도 수/인구 수

# 정보 중복 변수 삭제

예시)

ED	EE
robberies	robberPerPop
1	8.2
5	21.26
56	154.95
10	57.86
4	32.04

‘robberies’ : 강도 수

‘robberPerPop’ : 강도 수/인구 수

# 정보 중복 변수 삭제

rapes	rapesPerPop	robberies	robberiesPerPop	assaults	assaultsPerPop	burglaries	burglariesPerPop	larcenies	larceniesPerPop	autoTheft	autoTheftPerPop	arsons	arsonsPerPop
0	0	1	8.2	4	32.81	14	114.85	138	1132.08	16	131.26	2	16.41
1	4.25	5	21.26	24	102.05	57	242.37	376	1598.78	26	110.55	1	4.25
6	16.6	56	154.95	14	38.74	274	758.14	1797	4972.19	136	376.3	22	60.87
10	57.86	10	57.86	33	190.93	225	1301.78	716	4142.56	47	271.93	?	?
?	?	4	32.04	14	112.14	91	728.93	1060	8490.87	91	728.93	5	40.05
77	50.98	136	90.05	449	297.29	2094	1386.46	7690	5091.64	454	300.6	134	88.72
4	13.53	9	30.44	54	182.66	110	372.09	288	974.19	144	487.1	17	57.5
34	55.79	98	160.8	128	210.02	608	997.6	2250	3691.79	125	205.1	9	14.77
35	43.87	16	20.05	41	51.39	425	532.66	3149	3946.71	206	258.18	8	10.03
141	130.69	453	419.89	1043	966.77	2397	2221.81	6121	5673.63	1070	991.8	18	16.68
29	90.25	71	220.96	131	407.69	468	1456.49	1817	5654.8	151	469.94	6	18.67
21	77.73	309	1143.81	362	1340	478	1769.39	1460	5404.4	430	1591.71	20	74.03
36	78.53	58	126.52	269	586.8	582	1269.58	1786	3895.99	148	322.85	9	19.63
59	100.51	261	444.61	531	904.55	1754	2987.92	2010	3424.02	517	880.7	46	78.36
?	?	0	0	9	65.82	67	489.98	283	2069.62	21	153.58	1	7.31
?	?	27	79.7	37	109.22	226	667.12	1618	4776.1	159	469.34	7	20.66
0	0	23	50.46	17	37.29	282	618.62	906	1987.5	232	508.94	7	15.36
30	16.72	355	197.91	277	154.42	1596	889.74	4501	2509.23	1447	806.68	73	40.7
2	13.24	10	66.18	5	33.09	99	655.15	414	2739.73	18	119.12	4	26.47
146	49.94	710	242.88	1396	477.55	3977	1360.48	11514	3938.78	2452	838.8	97	33.18

# 정보 중복 변수 삭제

rapes	rapesPerPop	robberies	robberiesPerPop	assaults	assaultsPerPop	burglaries	burglariesPerPop	larcenies	larceniesPerPop	autoTheft	autoTheftPerPop	arsons	arsonsPerPop
0	0	1	8.2	4	32.81	14	114.85	138	1132.08	16	131.26	2	16.41
1	4.25	5	21.26	24	102.05	57	242.37	376	1598.78	26	110.55	1	4.25
6	16.6	56	154.95	14	38.74	274	758.14	1797	4972.19	136	376.3	22	60.87
10	57.86	10	57.86	33	190.93	225	1301.78	716	4142.56	47	271.93	?	?
?	?	4	32.04	14	112.14	91	728.93	1060	8490.87	91	728.93	5	40.05
77	50.98	136	90.05	449	297.29	2094	1386.46	7690	5091.64	454	300.6	134	88.72
4	13.53	9	30.44	54	182.66	110	372.09	288	974.19	144	487.1	17	57.5
34	55.79	98	160.8	128	210.02	608	997.6	2250	3691.79	125	205.1	9	14.77
35	43.87	16	20.05	41	51.39	425	532.66	3149	3946.71	206	258.18	8	10.03
141	130.69	453	419.89	1043	966.77	2397	2221.81	6121	5673.63	1070	991.8	18	16.68
29	90.25	71	220.96	131	407.69	468	1456.49	1817	5654.8	151	469.94	6	18.67
21	77.73	309	1143.81	362	1340	478	1769.39	1460	5404.4	430	1591.71	20	74.03
36	78.53	58	126.52	269	586.8	582	1269.58	1786	3895.99	148	322.85	9	19.63
59	100.51	261	444.61	531	904.55	1754	2987.92	2010	3424.02	517	880.7	46	78.36
?	?	0	0	9	65.82	67	489.98	283	2069.62	21	153.58	1	7.31
?	?	27	79.7	37	109.22	226	667.12	1618	4776.1	159	469.34	7	20.66
0	0	23	50.46	17	37.29	282	618.62	906	1987.5	232	508.94	7	15.36
30	16.72	355	197.91	277	154.42	1596	889.74	4501	2509.23	1447	806.68	73	40.7
2	13.24	10	66.18	5	33.09	99	655.15	414	2739.73	18	119.12	4	26.47
146	49.94	710	242.88	1396	477.55	3977	1360.48	11514	3938.78	2452	838.8	97	33.18





NA 확인

```
> result.naomit
```

	variables	mean	median	max	min	NAs
27	OtherPerCap	9442.765	8186	137000	0	1
100	LemasSwornFT	499.198	173	25655	65	1872
101	LemasSwFTPerPop	246.491	196.01	3437.23	29.4	1872
102	LemasSwFTFieldOps	432.56	152	22496	14	1872
103	LemasSwFTFieldPerPop	210.845	170.27	3290.62	19.21	1872
104	LemasTotalReq	252404.988	90000	8328470	2100	1872
105	LemasTotReqPerPop	120651.719	91034.6	1926281.5	2704.8	1872
106	PolicReqPerOffic	523.658	443.2	2162.5	20.8	1872
107	PolicPerPop	246.494	196	3437.2	29.4	1872
108	RacialMatchCommPol	85.5	87.93	100	42.15	1872
109	PctPolicWhite	82.516	86.18	100	1.6	1872
110	PctPolicBlack	9.263	5	67.31	0	1872
111	PctPolicHisp	5.46	2.04	98.4	0	1872
112	PctPolicAsian	0.681	0	18.57	0	1872
113	PctPolicMinor	15.242	11.37	98.4	0	1872
114	OfficAssgnDrugUnits	26.289	12	1773	0	1872
115	NumKindsDrugsSeiz	8.816	9	15	1	1872
116	PolicAveOTWorked	119.114	98.7	634.7	0	1872
120	PolicCars	185.478	86	3187	20	1872
121	PolicOperBudg	32176019.344	11164110	1617293056	2380215	1872
122	LemasPctPolicOnPatr	87.131	89.58	99.94	10.85	1872
123	LemasGangUnitDeploy	4.286	5	10	0	1872
125	PolicBudgPerPop	153577.871	114582	2422367	15260.4	1872
128	rapes	28.046	7	2818	0	208
129	rapesPerPop	36.258	26.92	401.35	0	208
130	robberies	237.952	19	86001	0	1
131	robberPerPop	162.613	74.8	2264.13	0	1
132	assaults	326.528	56	62778	0	13
133	assaultPerPop	378.005	226.525	4932.5	0	13
134	burglaries	761.237	205	99207	2	3
135	burglPerPop	1033.43	822.715	11881.02	16.92	3
136	larcenies	2137.629	747	235132	10	3
137	larcPerPop	3372.979	3079.51	25910.55	77.86	3
138	autoTheft	516.693	75	112464	1	3
139	autoTheftPerPop	473.966	302.355	4968.59	6.55	3
140	arsons	30.908	5	5119	0	91
141	arsonsPerPop	32.154	21.08	436.37	0	91
142	ViolentCrimesPerPop	589.079	374.06	4877.06	0	221
143	nonViolPerPop	4908.242	4425.45	27119.76	116.79	97

데이터 수 2215 개 중 1872 개가 NA인 변수  
(85%)

데이터 수 2215 개 중 91~221 개가 NA인 변수  
(4%~10%)

```
> result.naomit
```

	variables	mean	median	max	min	NAs
27	OtherPerCap	9442.765	8186	137000	0	1
100	LemasSwornFT	499.198	173	25655	65	1872
101	LemasSwFTPerPop	246.491	196.01	3437.23	29.4	1872
102	LemasSwFTFieldOps	432.56	152	22496	14	1872
103	LemasSwFTFieldPerPop	210.845	170.27	3290.62	19.21	1872
104	LemasTotalReq	252404.988	90000	8328470	2100	1872
105	LemasTotReqPerPop	120651.719	91034.6	1926281.5	2704.8	1872
106	PolicReqPerOffic	523.658	443.2	2162.5	20.8	1872
107	PolicPerPop	246.494	196	3437.2	29.4	1872
108	RacialMatchCommPol	85.5	87.93	100	42.15	1872
109	PctPolicWhite	82.516	86.18	100	1.6	1872
110	PctPolicBlack	9.263	5	67.31	0	1872
111	PctPolicHisp	5.46	2.04	98.4	0	1872
112	PctPolicAsian	0.681	0	18.57	0	1872
113	PctPolicMinor	15.242	11.37	98.4	0	1872
114	OfficAssgnDrugUnits	26.289	12	1773	0	1872
115	NumKindsDrugsSeiz	8.816	9	15	1	1872
116	PolicAveOTWorked	119.114	98.7	634.7	0	1872
120	PolicCars	185.478	86	3187	20	1872
121	PolicOperBudg	32176019.344	11164110	1617293056	2380215	1872
122	LemasPctPolicOnPatr	87.131	89.58	99.94	10.85	1872
123	LemasGangUnitDeploy	4.286	5	10	0	1872
125	PolicBudgPerPop	153577.871	114582	2422367	15260.4	1872
128	rapes	28.046	7	2818	0	208
129	rapesPerPop	36.258	26.92	401.35	0	208
130	robberies	237.952	19	86001	0	1
131	robberPerPop	162.613	74.8	2264.13	0	1
132	assaults	326.528	56	62778	0	13
133	assaultPerPop	378.005	226.525	4932.5	0	13
134	burglaries	761.237	205	99207	2	3
135	burglPerPop	1033.43	822.715	11881.02	16.92	3
136	larcenies	2137.629	747	235132	10	3
137	larcPerPop	3372.979	3079.51	25910.55	77.86	3
138	autoTheft	516.693	75	112464	1	3
139	autoTheftPerPop	473.966	302.355	4968.59	6.55	3
140	arsons	30.908	5	5119	0	91
141	arsonsPerPop	32.154	21.08	436.37	0	91
142	ViolentCrimesPerPop	589.079	374.06	4877.06	0	221
143	nonViolPerPop	4908.242	4425.45	27119.76	116.79	97

데이터 수 2215 개 중 1872 개가 NA인 변수  
(85%)

->삭제!

데이터 수 2215 개 중 91~221 개가 NA인 변수  
(4%~10%)

->임putation!

3

CLUSTERING

# CLUSTERING

```
description.txt
-- 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)
-- householdsSize: 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)
-- racePctHispan: 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)
-- numUrban: 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)
-- pctWage: 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)
```

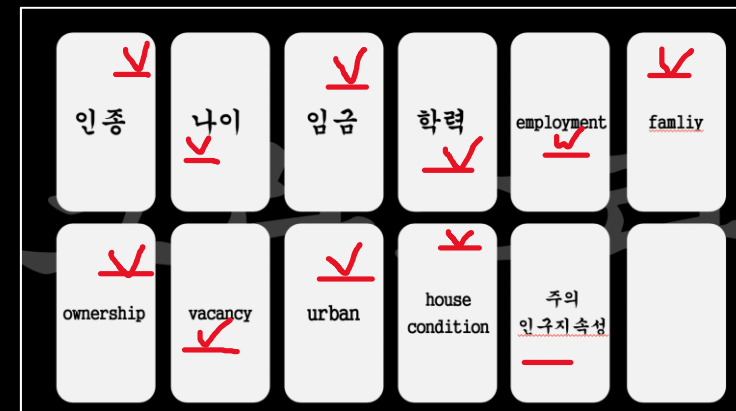
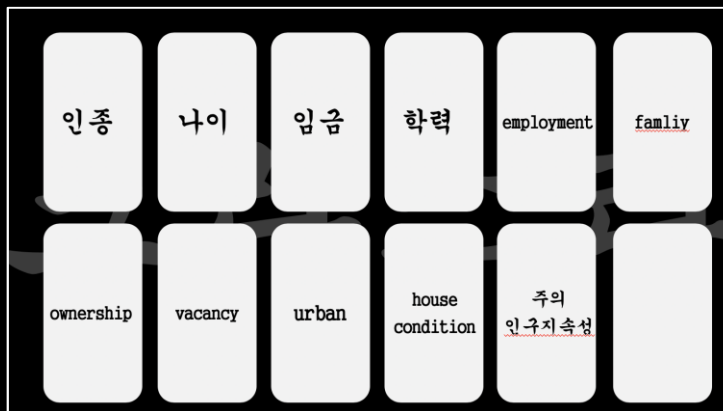
변수 디스트림션



군집화



대표 변수 선택



(주의) 한계점이 있습니다.

인종

나이

임금

학력

employment

famliy

ownership

vacancy

urban

house  
condition

주의  
인구지속성



racePctBlack  
racePctWhite  
racePctAsian  
racePctHispanic

agePct12t21  
agePct12t29  
agePct16t24  
agePct65up

medIncome  
pctWWage  
pctWFarmSelf  
pctWInvInc  
pctWSocSec  
pctWPubAss  
...

PctLess9thGrad  
PctNotHSGrad  
PctBSorMore

PctUnemployed  
PctEmploy  
PctEmplManu  
PctEmplProfServ  
PctOccupManu  
...

PctFam2Par  
PctKids2Par  
PctYoungKids2Par  
PctTeen2Par  
PctWorkMom  
...

OwnOccMedVal  
OwnOccHiQuart  
OwnOccQrange  
RentLowQ  
RentMedian  
RentHighQ

HousVacant  
PctHousOccup  
PctHousOwnOcc  
PctVacantBoarded  
PctVacMore  
6Mos

pctUrban  
NumStreet  
NumInShelters

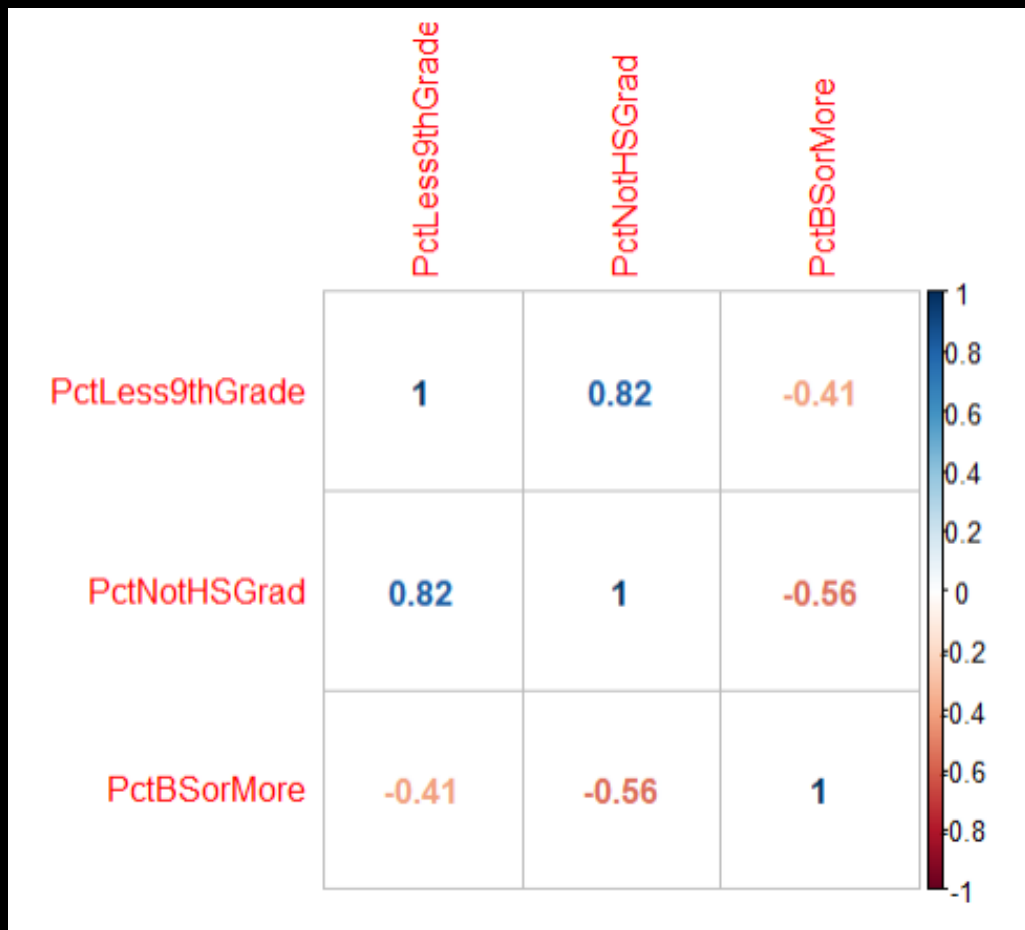
PctHousLess3BR  
PctPersDenseHous  
MedYrHousBuilt  
PctHousNoPhone  
PctW  
OFullPlumb

PctForeignBorn  
PctBorn-  
SameState  
PctSameHouse85  
PctSameState85

# 학력

변수

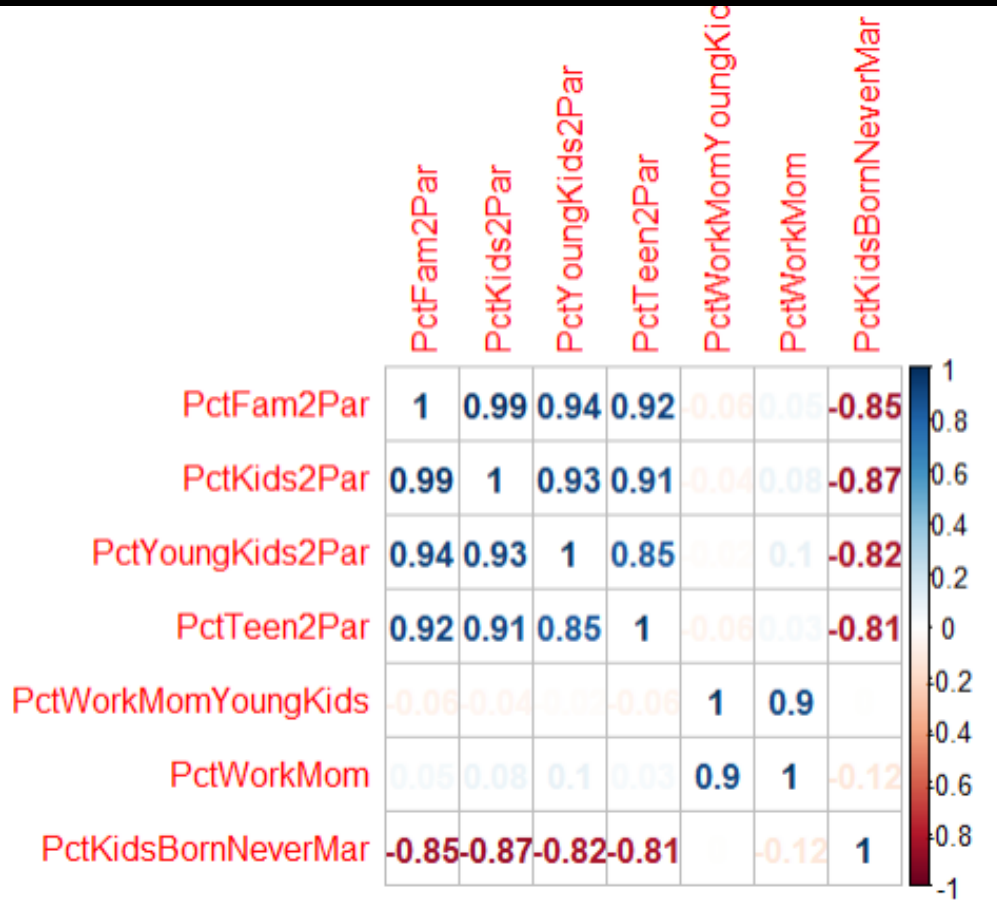
: 'racepctblack' , ' racePctWhite' ,  
'racePctAsian' , 'racePctHisp'



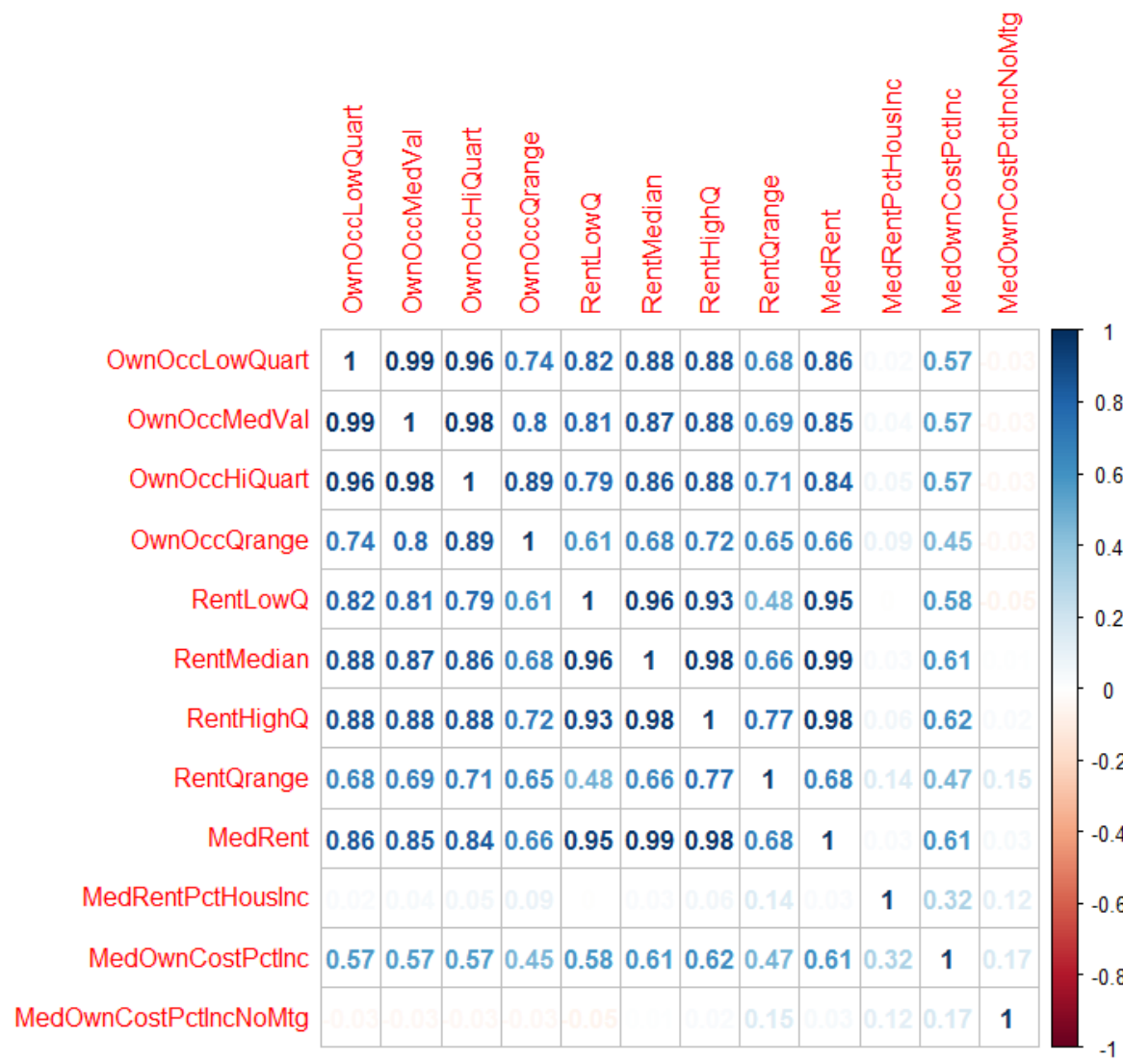
## family

변수

: 'PctFam2Par' , 'PctKids2Par' ,  
 'PctYoungKids2Par' , 'PctTeen2Par' ,  
 'PctWorkingMomYoungKids' ,  
 'PctWorkMom' , 'PctKidsBornNeverMar'



# ownership



변수

: 'OwnOccLowQuart', 'OwnOccMedVal',  
 'OwnOccHiQuart', 'OwnOccQrange',  
 'RentLowQ', 'RentMedian', 'RentHighQ',  
 'RentQrange', 'MedRent',  
 'MedRentPctHousInc', 'MedOwnCostPctInc',  
 'MedOwnCostPctIncNoMtg'

4

TRANSFORMATION

# 변수 변환 대상

” Skewed”

'pop', 'pctBlack', 'pctAsian', 'pctHisp', 'medIncome', 'pctWfarm', 'pctPubAsst', 'blackPerCap', 'NAper  
Cap',

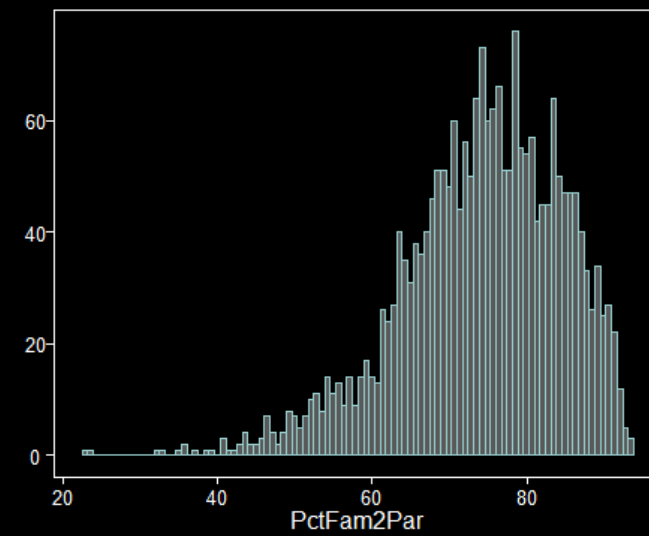
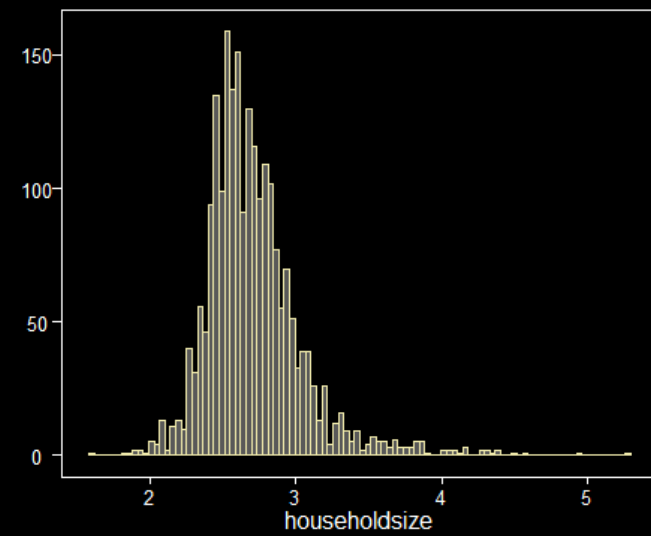
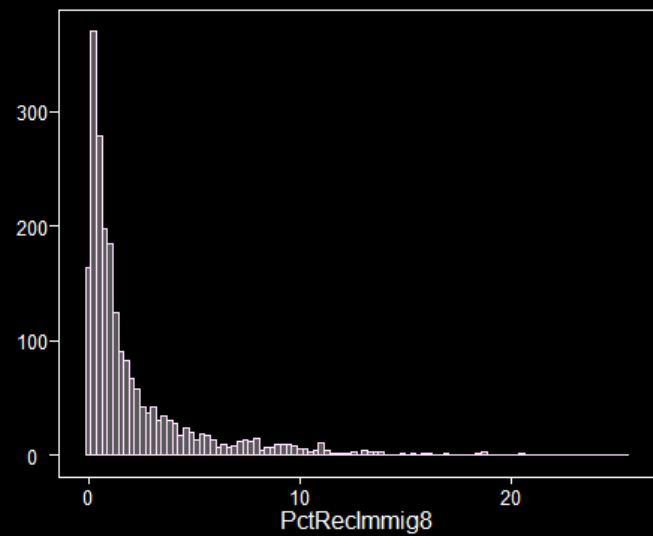
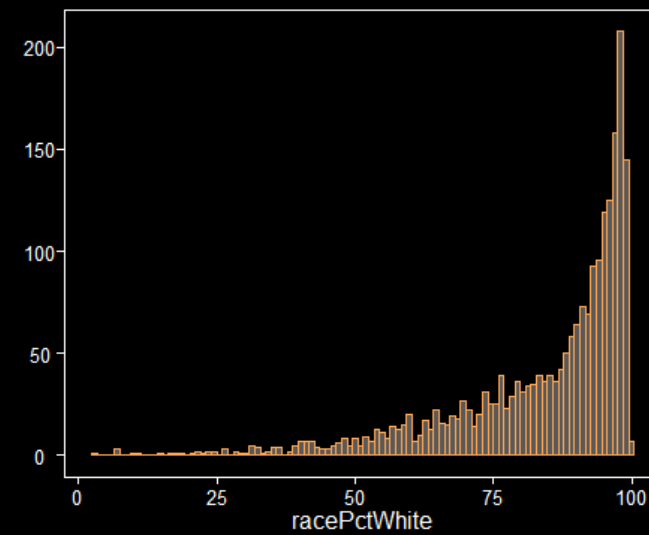
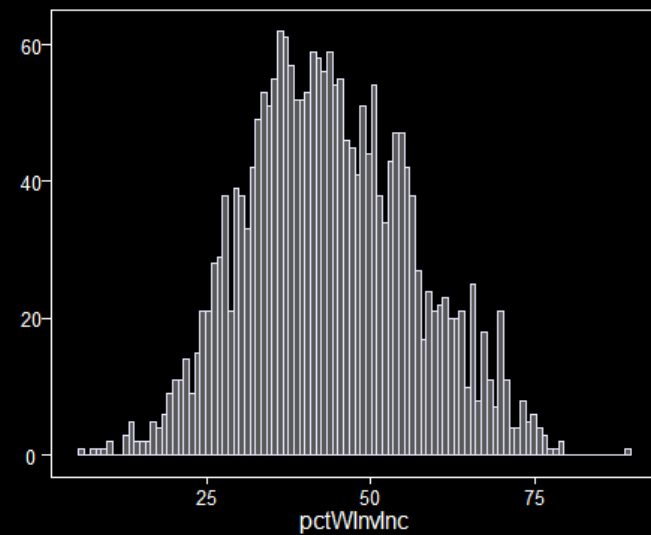
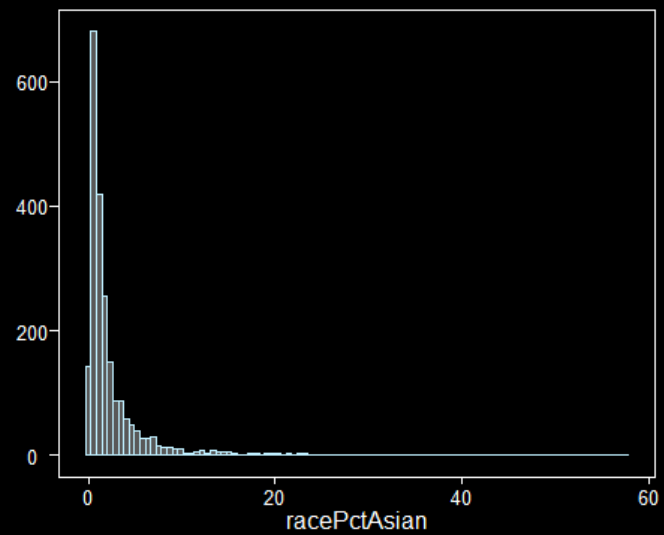
'asianPerCap', 'otherPerCap', 'hispPerCap', 'pctPoverty', 'pctLowEdu', 'pctKidsBornNevrMarr', 'numFo  
reignBorn',

'pctFgnImmig.3', 'pctFgnImmig.3.5', 'pctFgnImmig.5.8', 'pctFgnImmig.8.10', 'pctImmig.10', 'pctNotSp  
eakEng',

'pctLargHousFam', 'pctPopDenseHous', 'houseVacant', 'pctVacantBoarded', 'pctHousW0phone', 'pctHousW  
0plumb',

'ownHousMed', 'rentMed', 'medGrossRent', 'medOwnCostPctW0', 'persEmergShelt', 'pctForeignBorn', 'lan  
dArea',

'popDensity', 'pctUsePubTrans', 'pctOfficDrugUnit', 'nonviolentPerPop'



E

O

D

