

# INFO523 - Final Project

Dataset Notes | ELT JMP 2021-11-10

## Notes on Sample (Stata) and Subset (R)

- Sample taken in Stata from larger CVRM dataset (n=313,469)
- Sample size: n = 1000
- Sample is representative. See below [left = full set; right = sample]

## Race

race	Freq.	Percent	Cum.
API	3,603	1.15	1.15
BLK	168,885	53.88	55.03
I	227	0.07	55.10
U	1,574	0.50	55.60
WBH	1,671	0.53	56.13
WHI	63,485	20.25	76.39
WWH	74,024	23.61	100.00
Total	313,469	100.00	

race	Freq.	Percent	Cum.
API	15	1.50	1.50
BLK	530	53.00	54.50
I	2	0.20	54.70
U	4	0.40	55.10
WBH	1	0.10	55.20
WHI	223	22.30	77.50
WWH	225	22.50	100.00
Total	1,000	100.00	

## Risk Score

race	riskscore					Total
	HIGH	LOW	MODERATE	VERY HIGH	VERY LOW	
API	2	1,115	28	0	2,458	3,603
BLK	1,411	63,252	6,054	685	97,483	168,885
I	1	59	5	0	162	227
U	5	565	15	0	989	1,574
WBH	10	649	65	4	943	1,671
WHI	32	19,459	516	7	43,471	63,485
WWH	281	27,082	1,541	89	45,031	74,024
Total	1,742	112,181	8,224	785	190,537	313,469

race	riskscore					Total
	HIGH	LOW	MODERATE	VERY HIGH	VERY LOW	
API	0	5	0	0	10	15
BLK	2	219	18	1	290	530
I	0	0	0	0	2	2
U	0	3	0	0	1	4
WBH	0	0	0	0	1	1
WHI	0	69	1	0	153	223
WWH	1	76	7	0	141	225
Total	3	372	26	1	598	1,000

## City/County Residence

Lives in City or County	Freq.	Percent	Cum.
Rural/County	464	46.40	46.40
Chicago City	536	53.60	100.00
Total	1,000	100.00	

Lives in City or County	Freq.	Percent	Cum.
Rural/County	463	46.30	46.30
Chicago City	537	53.70	100.00
Total	1,000	100.00	

## Variables in Subset (9 vars total)

- Current age, days since last arrest, lives in City or county, race, sex, police district of residence, live where arrested, risk score, arrested by CPD