Caleb Lich, Jacob Blovins, Nathan Hynek, Michael Steets

9/2/2021

CSC 3220

1. What are the missing values
   a. Assault in Georgia is missing. A way to fix it would be to take the relation to murder to assault and do an average of it to Georgia, however, the values fluctuate between assault and murder so our number would not be right. As a second way to fix it, we could get rid of the Georgia data all together and deal with it later.
2. Look for outliers and smooth noisy data
   a. Iowa has 570 percent Urban population. We'd assume that the 0 was an accident and we make it 57%
   b. New York has a 6 percent urban population which we'd believe is false since most of the other areas are double digits. To fix it, we could generally google the urban population of New York in 1973 and put it in.
   c. 879 assault cases per 100,000 in South Carolina seemed a bit higher than the rest.
   d. 0.8 murder cases in North Dakota seemed a bit small.
3. Prepare the dataset to establish a relation between an urban population category and a crime type.

	Murder	Assault	UrbanPop	UrbanPop_divided by cell
New York	11.1	254	6	6
Vermont	2.2	48	32	32
West Virginia	5.7	81	39	39
Mississippi	16.1	259	44	44
North Dakota	0.8	45	44	44
North Carolina	13	337	45	45
South Dakota	3.8	86	45	45
South Carolina	14.4	879	48	48
Alaska	10	263	48	48
Arkansas	8.8	190	50	50
Maine	2.1	83	51	51
Kentucky	9.7	109	52	52
Montana	6	109	53	53
Idaho	2.6	120	54	54
New Hampshire	2.1	57	56	56
Alabama	13.2	236	58	58
Tennessee	13.2	188	59	59
Georgia	17.4		60	60
Wyoming	6.8	161	60	60
Nebraska	4.3	102	62	62
Virginia	8.5	156	63	63

Indiana	7.2	113	65	65
Louisiana	15.4	249	66	66
Kansas	6	115	66	66
Minnesota	2.7	72	66	66
Wisconsin	2.6	53	66	66
Maryland	11.3	300	67	67
Oregon	4.9	159	67	67
Oklahoma	6.6	151	68	68
New Mexico	11.4	285	70	70
Missouri	9	178	70	70
Pennsylvania	6.3	106	72	72
Delaware	5.9	238	72	72
Washington	4	145	73	73
Michigan	12.1	255	74	74
Ohio	7.3	120	75	75
Connecticut	3.3	110	77	77
Colorado	7.9	204	78	78
Florida	15.4	335	80	80
Texas	12.7	201	80	80
Arizona	8.1	294	80	80
Utah	3.2	120	80	80
Nevada	12.2	252	81	81
Illinois	10.4	249	83	83
Hawaii	5.3	46	83	83
Massachusetts	4.4	149	85	85
Rhode Island	3.4	174	87	87
New Jersey	7.4	159	89	89
California	9	276	91	91
Iowa	2.2	56	570	570

We did this so it's broken up more evenly instead of there being a lot of repeating numbers. We broke them up by 12, 12, 13, 13.