

Cluster Analysis of Crime Data in Tucson, Arizona

A. Carney

University of Arizona, Tucson, AZ,86721, USA

Crime data is being used by both law enforcement and policymakers to alleviate the complexity that comes with using public resources to combat crime and its precursors effectively. A wide variety of data mining techniques can be used on large quantities of public crime data to lend specific insights to concerned policymakers and public officials [1]. This historical cluster analysis of crime data could give policymakers insights into localities that need more (or less) focused policing and community resources. Cluster analysis can serve as a starting point for identifying crime hotspots in a given locality before performing any deeper analysis. An example of a deeper analysis stemming from intrigue provided by a cluster analysis is Exploratory Spatial Data Analysis (ESDA) [2]. An exploratory k-means cluster analysis of Tucson crime data is an excellent jumping off point for deeper analysis of criminality in Tucson. Using data provided by the Tucson Police Department via the Open Data initiative, it is possible to perform a cluster analysis by zip code to identify localities that are similar in their criminalities. Localities in Tucson can be clustered neatly into four distinct profiles of criminality. Merging with other datasets in the future could lend greater insight into what differentiates these localities.

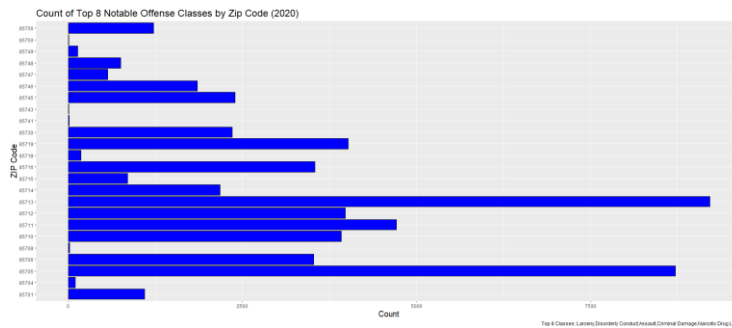


Fig.1 – Count of Top 8 notable Offense classes by zip code

ASSAULT	BURGLARY	CRIMINAL DAMAGE	DISORDERLY CONDUCT	FRAUD	GDV	LARCENY	NARCOTICS/DRUG LAWS	GROUP	ZIP
-0.3910865	-0.81939745	-0.37787768	-0.4841169	-0.433782	-0.3294417	-0.8870858	0.35274527	2	85701
-0.2619989	-0.39994028	-0.8852982	-0.7616363	-0.591413	-0.8935058	-0.2713339	-0.77233337	4	85704
1.9391285	2.352544791	2.0670337	2.2127219	2.2621939	2.0574894	2.73814366	2.99654734	3	85705
0.7671038	0.83061483	0.6304352	0.4586763	0.2482341	1.1688089	0.36093573	0.65203338	1	85706
-0.6993167	-0.32669452	-0.52781023	-0.8201847	-1.063938	-0.8445305	-0.91061193	-0.778917791	4	85708
0.8920246	0.915791404	0.73437942	0.4186398	1.4833337	0.8049962	0.57010735	0.48303423	1	85710
1.1465768	1.470582642	0.99515340	0.7636467	1.1253333	1.25824042	0.8862747	1.14124080	1	85711
0.6544485	0.87308789	1.1388828	0.8963023	0.6578142	0.31024472	0.33863385	0.268773811	1	85712
3.0040262	1.85842814	2.7777835	2.8518930	2.2142377	2.7891642	2.48941328	2.87182934	3	85713
-0.6868247	-0.21778774	-0.2521896	-0.8869739	-0.3019495	0.88311396	0.89911482	1.148871481	2	85714
-0.6314644	-0.53882388	-0.8118438	-0.64305426	-0.1867102	-0.71877228	-0.43262836	-0.819823777	4	85715
0.3933378	0.85212552	0.68840193	0.23887193	0.7514381	0.43889635	0.84733198	0.12039865	1	85716
-0.897373	-0.82504216	-0.83873364	-0.76665616	-1.0042019	-0.7782338	-0.31242099	-0.778917791	4	85718
0.4407814	1.08177233	0.8764108	0.74220720	0.6889172	0.52748831	0.72822228	0.303821871	1	85719
0.2894206	0.07643369	0.15076421	-0.07423563	0.4388806	-0.14407285	-0.17890680	0.04662052	2	85720
-0.8654207	-0.93887128	-0.83110862	-0.78650995	-1.1083862	-0.88483305	-0.81192115	-0.783032948	4	85741
-0.8693383	-0.32669452	-0.52781023	-0.8201847	-1.0742338	-0.8445305	-0.91844721	-0.783032948	4	85743
-0.0175403	0.020327344	-0.07140399	0.13027918	0.1194909	0.12281691	0.26473147	-0.295820394	2	85745
-0.4143173	-0.64971422	-0.47820211	-0.37999672	-0.5981187	-0.20205278	0.41962346	-0.301057148	2	85746
-0.4404681	-0.676293887	-0.63781172	-0.69393534	-0.2882184	-0.78814838	-0.72077610	-0.72996638	4	85747
-0.7132200	-0.85732940	-0.73289544	-0.63432870	-0.2198219	-0.78627283	-0.39214373	-0.69420784	4	85748
-0.8081439	-0.85689187	-0.88482831	-0.78833289	-0.8904326	-0.83527424	-0.86086185	-0.754457134	4	85749
-0.8674242	-0.82688452	-0.82781023	-0.80184286	-1.1083862	-0.88483305	-0.81192115	-0.783032948	4	85750
-0.41678049	-0.73438354	-0.47893503	-0.38279541	-0.433782	-0.1838882	-0.45831515	-0.326386368	2	85756

Fig.2 – Cluster overview and groupings by zip code.

References

- [1] Chen, Hsinchun, et al. "Crime data mining: an overview and case studies." *Proceedings of the 2003 annual national conference on Digital government research*. 2003.
- [2] Murray, Alan, et al. "Exploring Spatial Patterns of Crime Using Non-hierarchical Cluster Analysis." *National Science Foundation*. 2012.