

Tamara Glazer
Machine Learning and Public Policy
Homework 1 – Diagnostic Writeup

Problem 1 – Question 2 – Generate summary statistics for the crime reports data including but not limited to number of crimes of each type, how they change over time, and how they are different by neighborhood. Please use a combination of tables and graphs to present these summary stats.

Figure 1: Crimes by Type, 2017-2018

	2017	2018	total	percent_change
THEFT	64346	65081	129427	0.01
BATTERY	49214	49781	98995	0.01
CRIMINAL DAMAGE	29042	27806	56848	-0.04
ASSAULT	19302	20376	39678	0.06
DECEPTIVE PRACTICE	19028	18731	37759	-0.02
OTHER OFFENSE	17227	17130	34357	-0.01
BURGLARY	13001	11730	24731	-0.1
NARCOTICS	11658	12987	24645	0.11
ROBBERY	11877	9683	21560	-0.18
MOTOR VEHICLE THEFT	11406	9987	21393	-0.12
CRIMINAL TRESPASS	6812	6904	13716	0.01
WEAPONS VIOLATION	4686	5450	10136	0.16
OFFENSE INVOLVING CHILDREN	2269	2229	4498	-0.02
CRIM SEXUAL ASSAULT	1628	1626	3254	0
PUBLIC PEACE VIOLATION	1498	1369	2867	-0.09
INTERFERENCE WITH PUBLIC OFFICER	1086	1305	2391	0.2
SEX OFFENSE	1025	1104	2129	0.08
PROSTITUTION	735	718	1453	-0.02
HOMICIDE	676	586	1262	-0.13
ARSON	444	372	816	-0.16
LIQUOR LAW VIOLATION	191	268	459	0.4
GAMBLING	191	201	392	0.05
STALKING	188	203	391	0.08
KIDNAPPING	190	169	359	-0.11
INTIMIDATION	151	168	319	0.11
CONCEALED CARRY LICENSE VIOLATION	69	149	218	1.16
OBSCENITY	87	86	173	-0.01
NON-CRIMINAL	38	37	75	-0.03
PUBLIC INDECENCY	10	14	24	0.4
HUMAN TRAFFICKING	9	14	23	0.56
OTHER NARCOTIC VIOLATION	11	1	12	-0.91
NON-CRIMINAL (SUBJECT SPECIFIED)	2	3	5	0.5

Figure 2: Number of Crimes by Month, 2017-2018

		Number of Crimes by Month, 2017-2018																							
		2017-01	2017-02	2017-03	2017-04	2017-05	2017-06	2017-07	2017-08	2017-09	2017-10	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04	2018-05	2018-06	2018-07	2018-08	2018-09	2018-10	2018-11	2018-12
THEFT		4954	4395	4499	4802	5336	5790	6088	6460	5521	5732	5382	5387	4845	3890	4745	4993	5726	6012	6159	6517	5755	5643	5138	5658
BATTERY		3566	3413	3859	4101	4561	4650	4702	4421	4395	4167	3722	3657	3583	3199	3957	3799	4977	4638	4881	4613	4349	4297	3634	3854
RIMINAL DAMAGE		2424	2047	2238	2451	2518	2582	2745	2684	2414	2438	2335	2166	1928	1672	2337	2239	2669	2587	2725	2635	2287	2520	1964	2243
EPTIVE PRACTICE		1730	1495	1603	1515	1533	1557	1593	1598	1591	1534	1691	1539	1425	1643	1537	1648	1589	1685	1713	1483	1500	1417	1552	
OTHER OFFENSE		1564	1390	1520	1581	1633	1412	1530	1455	1258	1341	1287	1256	1324	1260	1467	1501	1597	1506	1562	1487	1306	1381	1414	1325
ASSAULT		1373	1295	1486	1638	1852	1857	1819	1740	1786	1628	1454	1374	1416	1353	1618	1639	2034	1877	1941	1846	1811	1714	1539	1588
BURGLARY		1274	1012	931	968	1061	1090	1206	1152	1041	1199	1083	984	950	698	822	942	973	1062	1139	1213	1051	1003	910	967
R VEHICLE THEFT		1174	879	801	837	832	1014	991	952	915	986	955	1070	1006	626	737	751	830	811	865	927	905	793	817	919
ROBBERY		1075	787	783	851	903	981	1004	1174	1002	1118	1128	1071	898	636	738	694	802	858	876	942	768	838	789	844
NARCOTICS		980	954	1126	1042	1116	863	1000	993	1027	844	924	789	1078	994	1277	1098	1172	1076	1136	1148	1034	958	1081	935
MINAL TRESPASS		573	527	548	548	631	616	621	597	539	572	533	507	577	548	586	586	599	590	596	598	606	562	501	555
IPONS VIOLATION		336	331	307	406	397	442	519	456	418	389	337	348	410	306	393	399	541	498	549	561	487	451	407	448
OLVING CHILDREN		298	123	179	174	206	185	167	194	185	184	167	207	237	158	186	186	201	195	168	174	181	206	170	167
SEXUAL ASSAULT		150	124	103	143	131	121	163	157	145	147	124	120	113	114	117	120	146	148	154	149	156	144	135	130
PEACE VIOLATION		123	99	126	129	153	126	148	145	131	129	105	84	89	78	119	115	124	137	116	121	111	111	104	144
SEX OFFENSE		94	71	79	77	92	91	100	77	93	77	92	82	70	65	86	83	122	84	87	123	112	112	80	80
PROSTITUTION		86	52	68	88	74	76	61	81	23	25	64	37	31	20	38	101	95	65	46	74	69	48	85	46
I PUBLIC OFFICER		84	85	81	98	105	97	107	104	86	89	72	78	86	93	119	106	123	115	132	127	116	103	88	97
HOMICIDE		55	50	42	49	57	87	76	58	61	60	41	40	42	42	33	43	53	57	62	58	53	54	46	43
ARSON		38	40	27	47	38	46	34	43	33	39	38	21	26	14	32	27	33	36	41	38	32	35	26	32
STALKING		17	18	19	24	13	14	20	16	19	11	10	7	15	13	15	17	24	18	26	15	12	15	15	18
R LAW VIOLATION		10	22	27	7	20	20	28	19	13	7	9	9	23	9	23	20	32	26	32	25	28	21	20	9
INTIMIDATION		9	15	14	11	8	15	13	16	16	11	9	14	5	9	13	8	14	14	24	19	14	19	14	15
KIDNAPPING		8	13	18	15	19	16	23	13	22	11	16	16	16	14	14	12	17	18	13	15	12	7	16	15
ENSE VIOLATION		7	3	4	5	6	3	6	8	6	9	6	6	4	6	9	11	15	14	13	17	20	19	9	12
OBSCENITY		3	4	10	5	8	6	6	4	8	20	6	7	10	9	10	2	8	3	13	7	9	7	2	6
NON-CRIMINAL		3	1	1	1	3	6	4	5	1	4	7	2	4	3	3	4	2	2	3	3	4	4	3	2
GAMBLING		1	9	4	21	20	30	21	34	23	17	9	2	5	1	8	10	23	34	42	39	18	13	6	2
JBLIC INDECENCY		1	1		1	3				1	1		2				1	2	4	1	1	1	1	2	1
JECT SPECIFIED)			1	1													1			1	1				
MAN TRAFFICKING				4		1	1		1		2			2	1	2		1	1		1	3	1		2
COTIC VIOLATION				1				2	2	1		5												1	

Figure 3: Running Total of Reported Cases by Arrest Status, 2017-2018

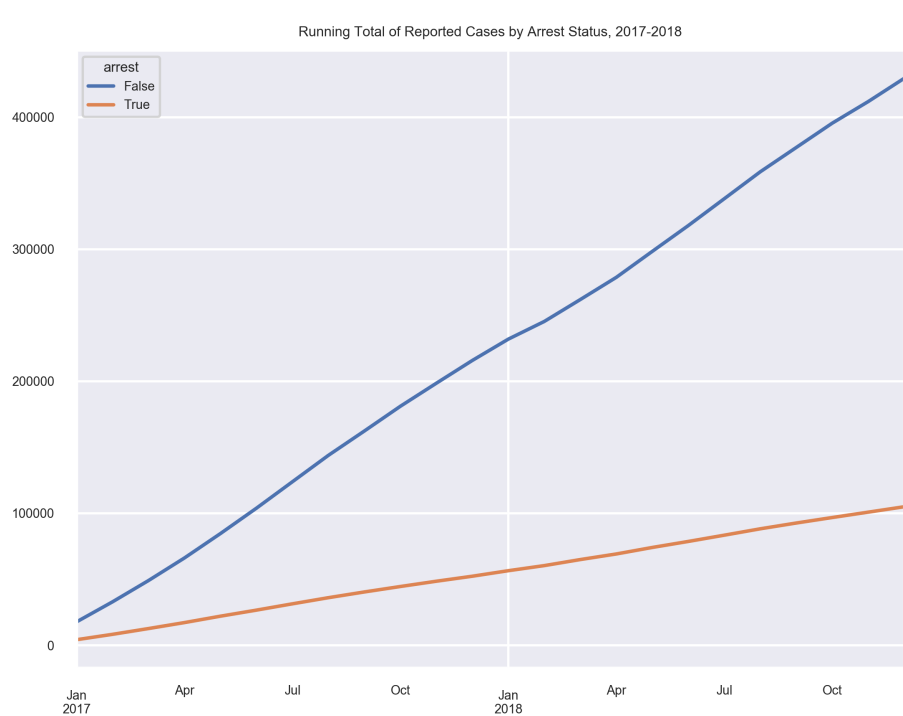


Figure 4: Top 5 Neighborhoods by Number of Reported Cases, 2017-2018

neighborhood	number_of_cases	number_of_arrests
Austin	30429	7794
Near North Side	25367	4240
Loop	21480	3239
Near West Side	18448	2643
North Lawndale	18377	6937

Figure 5: Bottom 5 Neighborhoods by Number of Reported Cases, 2017-2018

neighborhood	number_of_cases	number_of_arrests
Edison Park	528	58
Burnside	755	169
Forest Glen	1042	76
Mount Greenwood	1149	120
Montclare	1205	139

Figure 6: Crimes by Type and Neighborhood, 2017-2018

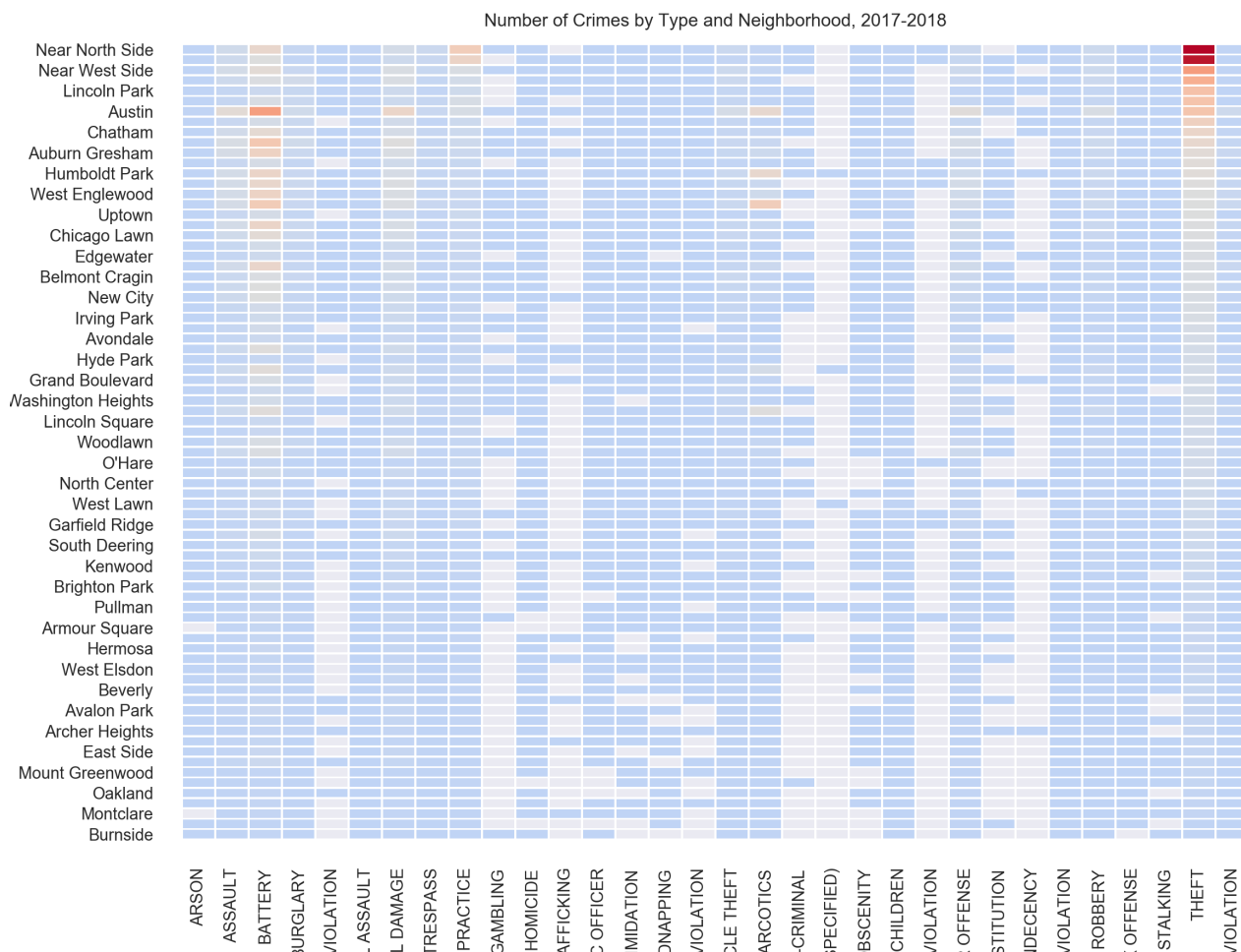


Figure 7: Number of Reported Crimes and Number of Arrests per Month, 2017-2018

month	number_of_cases	number_of_arrests
2017-01	22010	4265
2017-02	19256	4067
2017-03	20509	4380
2017-04	21635	4520
2017-05	23330	4853
2017-06	23794	4584
2017-07	24797	4772
2017-08	24659	4651
2017-09	22771	4340
2017-10	22848	4125
2017-11	21454	3999
2017-12	21034	3790
2018-01	20332	4182
2018-02	17256	3861
2018-03	21147	4574
2018-04	21045	4233
2018-05	24603	4934
2018-06	24075	4548
2018-07	25088	4746
2018-08	25207	4826
2018-09	22793	4380
2018-10	22580	4164
2018-11	20433	4070
2018-12	21709	4093

Problem 2 – Question 1 – What types of blocks have reports of “Battery”?

Figure 8: Chicago Statistics by Zip Code (ACS 2017-2018)

	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.18	44829.48	0.29
min	0.03	22198	0.01
25%	0.09	29207.25	0.04
50%	0.17	37114	0.13
75%	0.25	54994.5	0.52
max	0.44	113056	0.96

Figure 9: Battery Location Statistics by Incident Zip Code

	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.25	33627.74	0.5
min	0.03	22198	0
25%	0.16	25609	0.13
50%	0.25	28030	0.53
75%	0.33	37038	0.92
max	0.44	113056	0.96

According to 2017-2018 Crime Reports from the Chicago Open Data Portal joined with ACS data at the zip code level, the blocks/zip codes that have reports of battery have the following characteristics:

- On average, 25% of residents in these locations live below the poverty line as compared to an average of 18% across Chicago (**7% higher than Chicago average**). At the 25th percentile, 16% of residents live below the poverty line and at the 75th percentile, 33% of residents live below the poverty line.
- On average, the median annual earnings in these locations is \$33,627.74 as compared to an average median annual earnings of \$44,829.48 across Chicago (**\$11,201.74 lower than Chicago average**). At the 25th percentile, median annual earnings is \$22,198 and at the 75th percentile, median annual earnings is \$37,038, which is still lower than the average for Chicago.
- On average, 50% of residents in these locations are black as compared to an average of 29% across Chicago (**21% higher than Chicago average**). At the 25th percentile, 13% of residents are black and at the 75th percentile, 92% of residents are black, as compared to a maximum of 96% for Chicago.

Problem 2 – Question 2 – What types of blocks have reports of “Homicide”?

Figure 10: Chicago Statistics by Zip Code (ACS 2017-2018)

	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.18	44829.48	0.29
min	0.03	22198	0.01
25%	0.09	29207.25	0.04
50%	0.17	37114	0.13
75%	0.25	54994.5	0.52
max	0.44	113056	0.96

Figure 11: Homicide Location Statistics by Incident Zip Code

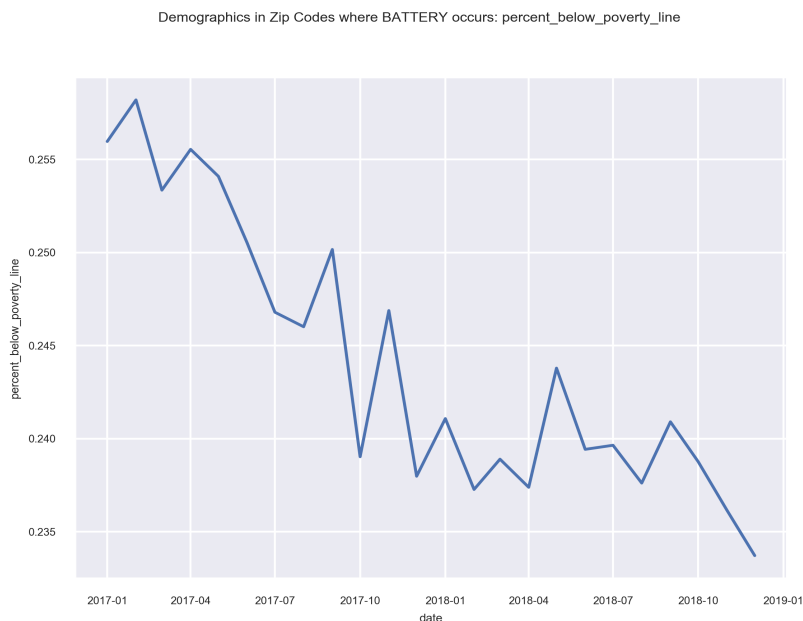
	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.28	30446.59	0.63
min	0.03	22198	0.01
25%	0.23	25151	0.24
50%	0.27	27703	0.77
75%	0.34	30396	0.93
max	0.44	95956	0.96

According to 2017-2018 Crime Reports from the Chicago Open Data Portal joined with ACS data at the zip code level, the blocks/zip codes that have reports of homicide have the following characteristics:

- On average, 28% of residents in these locations live below the poverty line as compared to an average of 18% across Chicago (**10% higher than Chicago average and 3% higher than battery location average**). At the 25th percentile, 23% of residents live below the poverty line and at the 75th percentile, 34% of residents live below the poverty line.
- On average, the median annual earnings in these locations is \$30,446.59 as compared to an average median annual earnings of \$44,829.48 across Chicago (**\$14,382.89 lower than Chicago average and \$3,181.15 lower than battery location average**). At the 25th percentile, median annual earnings is \$25,151 and at the 75th percentile, median annual earnings is \$30,396, which is still lower than the average for Chicago.
- On average, 63% of residents in these locations are black as compared to an average of 29% across Chicago (**34% higher than Chicago average and 113% higher than battery location average**). At the 25th percentile, 24% of residents are black and at the 75th percentile, 93% of residents are black, as compared to a maximum of 96% for Chicago.

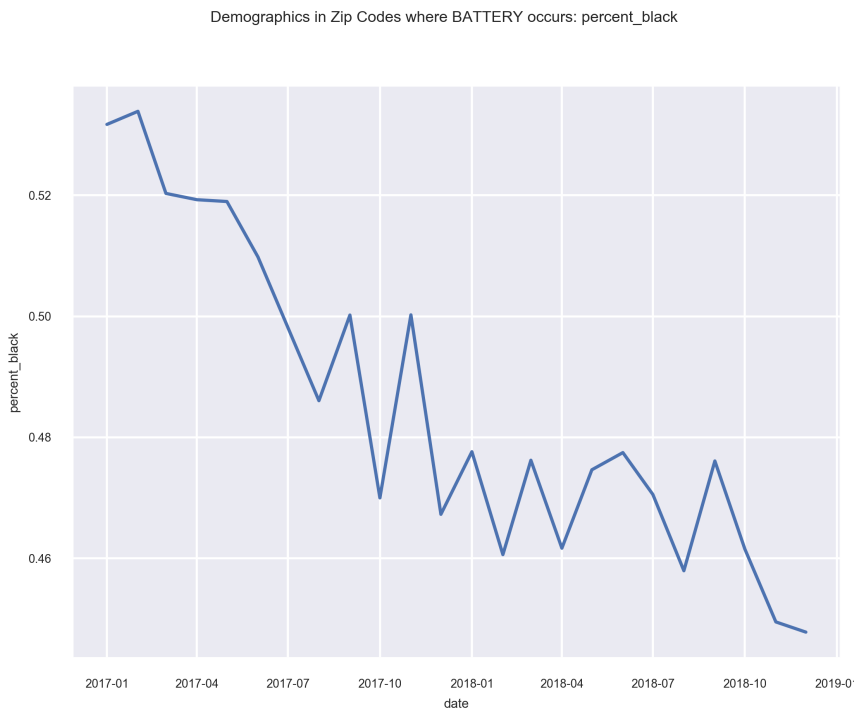
Problem 2 – Question 3 – Does that change over time in the data you collected?

Figure 12: Demographics in Zip Codes where Battery Occurs: Percent below Poverty Line



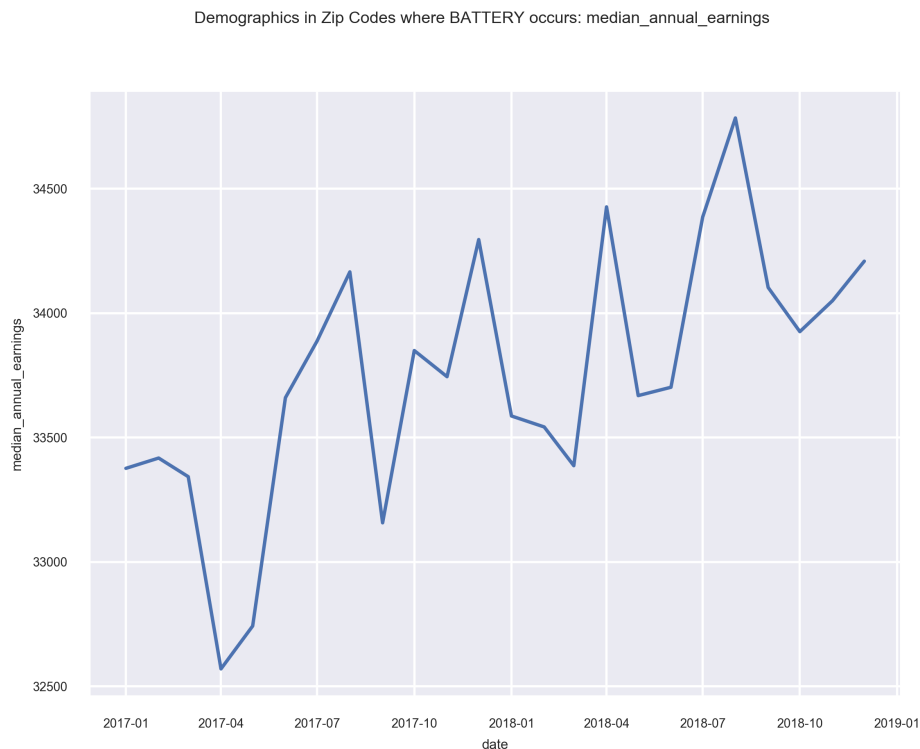
From January 2017 to December 2018, the average percentage of people below the poverty line in zip codes where battery occurs **decreases**.

Figure 13: Demographics in Zip Codes where Battery Occurs: Percent Black



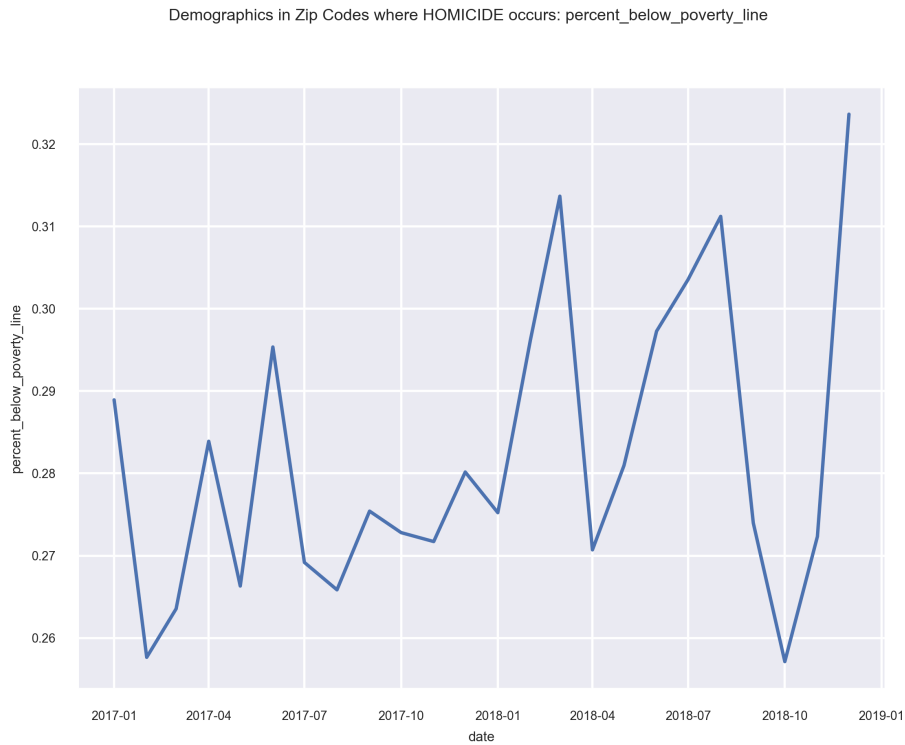
From January 2017 to December 2018, the average percentage of people who are black in zip codes where battery occurs **decreases**.

Figure 14: Demographics in Zip Codes where Battery Occurs: Median Annual Earnings



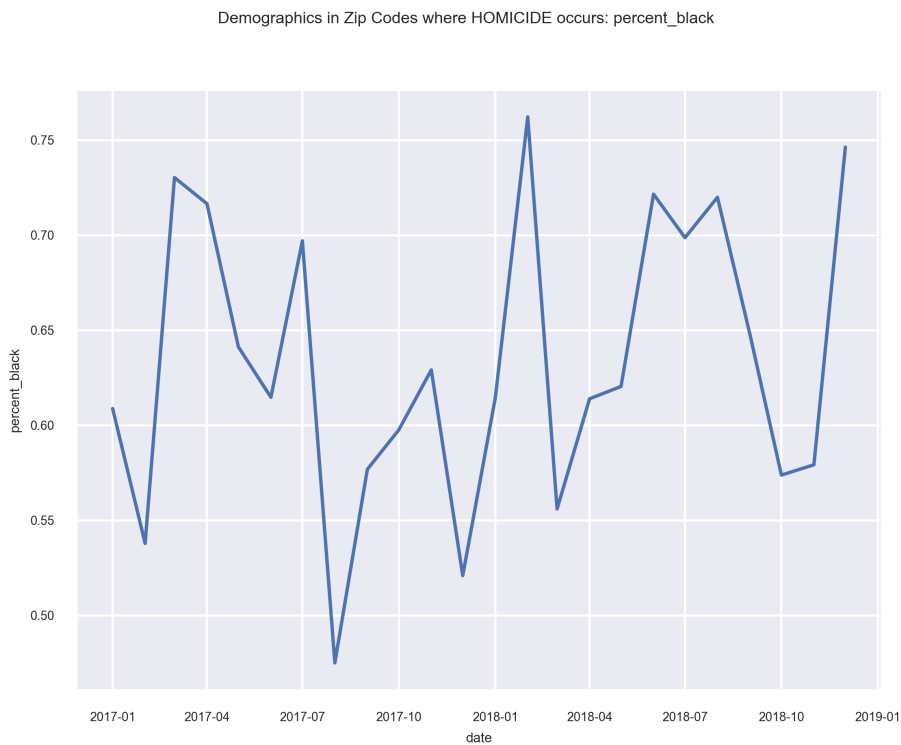
From January 2017 to December 2018, the average median annual earnings in zip codes where battery occurs **increases**.

Figure 15: Demographics in Zip Codes where Homicide Occurs: Percent Below Poverty Line



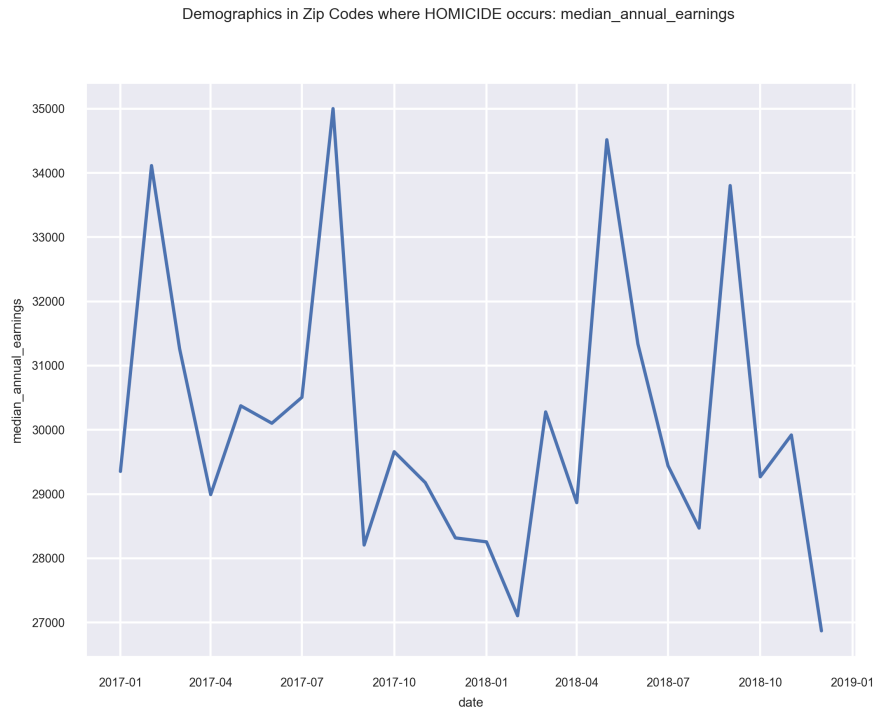
From January 2017 to December 2018, the average percentage of people below the poverty line in zip codes where homicide occurs **varies over time**, with signs of seasonality. Overall, the **trend is positive**.

Figure 16: Demographics in Zip Codes where Homicide Occurs: Percent Black



From January 2017 to December 2018, the average percentage of people that are black in zip codes where homicide occurs **varies inconsistently over time**, with signs of seasonality. The trend is not noticeably positive or negative.

Figure 17: Demographics in Zip Codes where Homicide Occurs: Median Annual Earnings



From January 2017 to December 2018, the average median annual earnings in zip codes where homicide occurs **varies inconsistently over time**, with signs of seasonality. The trend is **slightly negative**.

Problem 2 – Question 4 – What is the difference in blocks that get “Deceptive Practice” vs “Sex Offense”?

Figure 18: Deceptive Practice Location Statistics by Incident Zip Code

	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.2	39921.68	0.34
min	0.03	22198	0
25%	0.12	27933	0.04
50%	0.2	35986	0.17
75%	0.26	42154	0.73
max	0.44	113056	0.96

Figure 19: Sex Offense Location Statistics by Incident Zip Code

	percent_below_poverty_line	median_annual_earnings	percent_black
mean	0.22	36231.07	0.35
min	0.03	22198	0
25%	0.13	26161	0.05
50%	0.2	31212	0.18
75%	0.27	41268	0.73
max	0.44	113056	0.96

According to 2017-2018 Crime Reports from the Chicago Open Data Portal joined with ACS data at the zip code level, the blocks/zip codes that have reports of deceptive practice share similar characteristics with blocks/zip codes that have reports of sex offense. Within zip codes with reports of deceptive practice, 20% of

residents, on average, live below the poverty line, as compared to 22% of residents in zip codes with reports of sexual offense. The range of values is also very similar, with a 25th percentile of 12% and 13% in locations with deceptive practice and sexual offense, respectively, and a 75th percentile of 26% and 27%, respectively. These summary statistics indicate a slightly higher percentage of people below the poverty line in locations where sex offenses have been reported. Further analysis would need to be conducted to determine if this difference is at a statistically significant level. This is as compared to an average of 18% across Chicago.

Similarly, within zip codes with reports of deceptive practice, 34% of residents, on average, live below the poverty line, as compared to 35% of residents in zip codes with reports of sexual offense. The range of values is also very similar, with a 25th percentile of 4% and 5% in locations with deceptive practice and sexual offense, respectively, and a 75th percentile of 73% in both locations. These summary statistics indicate that a slightly higher percentage of people are black in locations where sex offenses have been reported. Further analysis would need to be conducted to determine if this difference is at a statistically significant level. This is as compared to an average of 29% across Chicago.

However, there is an economically significant difference in median annual earnings between locations where deceptive practice and sexual offense occur. Within zip codes with reports of deceptive practice, the average median income is \$39,921.68, which is \$3,690.61 higher than zip codes with reports of sexual offense (median income of \$36,231.07). Both values are compared to an average median income of \$44,829.48 across the greater Chicago area. The interquartile range is similarly shifted downward in locations where sex offenses are reported, with a 25th percentile of \$27,933 and a 75th percentile of \$42,154 in locations with reported deceptive practice, and respective values of \$26,161 and \$41,268 in locations with reported sexual offenses.

Problem 3 – Question 1 – Describe how crime has changed in Chicago from 2017 to 2018?

Between 2017 and 2018, reported crimes have decreased in absolute numbers by 1,829 or 0.68% (268,097 to 266,268), assuming consistent population size estimates between the two years (2018 population metrics not available). Meanwhile, total number of arrests have slightly increased in absolute numbers by 265 or 0.51% (52,346 to 52,611), and percent of crimes that result in arrest increased very slightly from 19.52% to 19.76%. Theft remains the highest type of reported crime and increased slightly from 64,346 to 65,081 between the two years (figure 1), representing an increase of 1%. It saw its highest peak in August 2018 at 6517 followed by August 2017 at 6,460, demonstrating seasonality. The crime type that saw the largest percent increase in occurrence was Concealed Carry License Violation (116% increase) and the crime type that saw the largest decrease in occurrence was Other Narcotic Violation (91% decrease).

Furthermore, the types of block/zip codes that have reports of battery changed in significant economic terms from 2017 to 2018. The average percentage of people below the poverty line in zip codes where battery occurs decreased relatively consistently from 25.1% to 23.9% from 2017 to 2018, which may signify a shift in crime to more affluent neighborhoods or an improvement in the financial situation within places that crime is already taking place. The average percentage of residents who are black in zip codes where battery occurs decreased relatively consistently from 50.9% to 46.7% from 2017 to 2018, and the average median annual earnings in zip codes where battery occurs increased from \$33,445 to \$33,982. The opposite trends were seen in locations where homicide occurred, with wider monthly variation (27.5% to 28.8% in percentage of people below the poverty line, 60.9% to 65.5% in percentage of people who are black, and \$30,601 to \$30,225 decrease in median earnings).

Problem 3 – Question 2a – Are these statistics correct?

Figure 20: Crimes by Type in Chicago, One Month [June 28th – July 25th]

	2017	2018	percent_change
ARSON	33	34	0.03
ASSAULT	1605	1723	0.07
BATTERY	4033	4352	0.08
BURGLARY	1060	997	-0.06
CONCEALED CARRY LICENSE VIOLATION	5	11	1.2
CRIM SEXUAL ASSAULT	129	132	0.02
CRIMINAL DAMAGE	2348	2332	-0.01
CRIMINAL TRESPASS	544	496	-0.09
DECEPTIVE PRACTICE	1385	1464	0.06
GAMBLING	14	38	1.71
HOMICIDE	70	49	-0.3
INTERFERENCE WITH PUBLIC OFFICER	94	118	0.26
INTIMIDATION	13	16	0.23
KIDNAPPING	18	10	-0.44
LIQUOR LAW VIOLATION	23	30	0.3
MOTOR VEHICLE THEFT	877	743	-0.15
NARCOTICS	871	1016	0.17
NON-CRIMINAL	3	3	0
NON-CRIMINAL (SUBJECT SPECIFIED)	0	1	inf
OBSCENITY	7	5	-0.29
OFFENSE INVOLVING CHILDREN	146	156	0.07
OTHER NARCOTIC VIOLATION	2	0	-1
OTHER OFFENSE	1307	1355	0.04
PROSTITUTION	64	33	-0.48
PUBLIC INDECENCY	0	1	inf
PUBLIC PEACE VIOLATION	128	107	-0.16
ROBBERY	874	795	-0.09
SEX OFFENSE	85	80	-0.06
STALKING	17	21	0.24
THEFT	5224	5378	0.03
WEAPONS VIOLATION	445	480	0.08
TOTAL	21424	21976	0.03

The 28 days leading up to July 26, 2018 (June 28, 2018 to July 25, 2018) as compared to the same 28 days in 2017 demonstrate the following:

- 9% decrease in Robberies (not a 21% increase)
- 8% increase in Batteries (not a 136% increase)
- 6% decrease in Burglaries (not an increase of 50%)

- 15% decrease in Motor Vehicle Theft (not an increase of 41%)
- 3% increase Total (not an increase of 16%)

Therefore, the statistics provided by the alderman candidate are not correct based on this dataset.

Figure 21: Crimes by Type in Chicago, Year to Date [January 1st – July 25th]

	2017	2018	percent_change
ARSON	270	209	-0.23
ASSAULT	11320	11878	0.05
BATTERY	28852	29034	0.01
BURGLARY	7542	6586	-0.13
CONCEALED CARRY LICENSE VIOLATION	34	72	1.12
CRIM SEXUAL ASSAULT	935	912	-0.02
CRIMINAL DAMAGE	17005	16157	-0.05
CRIMINAL TRESPASS	4064	4082	0
DECEPTIVE PRACTICE	11026	11066	0
GAMBLING	106	123	0.16
HOMICIDE	416	332	-0.2
HUMAN TRAFFICKING	6	7	0.17
INTERFERENCE WITH PUBLIC OFFICER	657	774	0.18
INTIMIDATION	85	87	0.02
KIDNAPPING	112	104	-0.07
LIQUOR LAW VIOLATION	134	165	0.23
MOTOR VEHICLE THEFT	6528	5626	-0.14
NARCOTICS	7081	7831	0.11
NON-CRIMINAL	19	21	0.11
NON-CRIMINAL (SUBJECT SPECIFIED)	2	2	0
OBSCENITY	42	55	0.31
OFFENSE INVOLVING CHILDREN	1332	1331	0
OTHER NARCOTIC VIOLATION	3	0	-1
OTHER OFFENSE	10630	10217	-0.04
PROSTITUTION	505	396	-0.22
PUBLIC INDECENCY	6	8	0.33
PUBLIC PEACE VIOLATION	904	778	-0.14
ROBBERY	6384	5502	-0.14
SEX OFFENSE	604	597	-0.01
STALKING	125	128	0.02
THEFT	35864	36370	0.01
WEAPONS VIOLATION	2738	3096	0.13
TOTAL	155331	153546	-0.01

Year-to-date leading up to July 26, 2018 (January 1, 2018 to July 25, 2018) as compared to the same period in 2017 demonstrates a 1% decrease in total reported crime from 2017 to 2018 (not an increase of 10% since

2017). Therefore, the year-to-date statistics provided by the alderman candidate are not correct based on this dataset.

Problem 3 – Question 2b – Could they be misleading or would you agree with the conclusions he’s drawing? Why or why not?

I believe that the alderman candidate’s statistics are misleading and a bit exaggerated. According to Crime Reports from the Open Data Portal for 2017 to 2018, crime increased by 3% during the one-month period noted and decreased by 1% year-to-date since 2017. Crimes are not raising at the rates he proposes. That said, it is possible that he is referring to metrics specific to the 43rd Ward, which is not indicated directly on his website, and I do generally agree with the conclusions he is drawing around crime being an important issue in our neighborhoods. Crime rates have been relatively consistent over the past two years, so increased police presence in places that experience the highest crime rates may help to reduce crime in significant ways. I also agree that officers should know the names of people in their communities. However, the data generally indicate positive trends in crime rates and I believe his solution should address public education and provision of social services instead of strictly increasing police presence since the figures he draws upon are inaccurate.

Problem 3 – Question 3 – As you know, there will be a new mayor in Chicago very soon. Based on these summary statistics, provide 5 key findings to the new mayor’s office about crime in Chicago and what they should focus on in order to deal with crime in Chicago.

Based on the summary statistics calculated above, I would provide the following 5 key findings and recommendations to the new Mayor’s office about crime in Chicago:

1. While reported crime has decreased in absolute terms since 2017, theft and battery are consistently the two most prevalent types of crime in Chicago making up a cumulative 43% of all reports in 2018, each having increased by 1% over the past year.

Recommendation: Focus on addressing the root causes of theft and battery across the Chicago metropolitan area, including increasing public awareness and education around alternative social services available to individuals in need.

2. Crime in Chicago appears to be seasonal and cyclical in nature, with noticeable spikes between July and August each year as well as a noticeable drop in February.

Recommendation: Allocate time and resources to better research and understand the seasonal nature of crime in Chicago. Consider increasing police presence in high risk neighborhoods during summer and winter months in line with historical spikes in criminal activity.

3. The top 5 neighborhoods in Chicago with the most crime are Austin, Near North Side, the Loop, Near West Side, and North Lawndale.

Recommendation: Increase social service provision in these key high-risk neighborhoods, including increasing the number of shelters available during winter months and increasing the accessibility of existing social services like Food Stamps (eg. making forms easier to fill out and submit), in order to reduce petty theft.

4. While battery incidents in absolute terms have increased since 2017, the places where they are occurring tend to have lower percentages of people below the poverty line over time, which may indicate a shift in where these crimes are occurring.

Recommendation: Continue to research shifts in where crimes are taking place over time, in order to better anticipate where future incidents will occur so that you can increase police presence and public safety awareness in these areas at appropriate times.

5. Zip codes that tend to experience the highest crime rates in Chicago typically have a higher percentage of residents below the poverty line and have a lower median income than averages for the greater Chicago area.

Recommendation: Focus on addressing the root causes of poverty and low earnings in high-risk neighborhoods to mitigate future crime, including expanding job programs, promoting job creation, and improving low-cost educational opportunities.

Problem 3 – Question 4 – What are some of the key caveats of your recommendations and limitations of the analysis that you just did?

There are several limitations to this analysis that should be considered when assessing the credibility of findings and action to take. First, because only two years of crime data are used, it is difficult to track long-term crime trends and progress towards crime reduction. It is possible, for instance, that 2018 could be an outlier over a long time period, which would be difficult to capture with only two years of data. Additionally, this analysis relies on trends rather than a rigorous regression. As a result, I am able to highlight correlation between low earnings, poverty rates, race and crime rates, but cannot claim that these relationships are causal. It is possible that the variables selected are proxies for the true mechanisms causing variation in crime by location, which cannot be determined without leveraging causal inference statistics. Furthermore, while this analysis is conducted at the zip code level, it is possible that there are important differences in crime prevalence at the street level. Analysis at a more granular level could help to inform neighborhood specific recommendations to mitigate future crime. Finally, while this analysis is limited to a few specific variables of interest from the ACS, it would be interesting to consider how other metrics correlate with crime prevalence, including education levels, family size, and geographic proximity to police stations.

There are additional caveats to my recommendations that must be considered. Specifically, my recommendations assume that crimes are being committed largely by individuals who live in the zip codes in which the crimes are reported, while it is possible that people travel across zip codes to commit crimes. Such patterns must be investigated to better understand where to direct social services as well as public safety campaigns. Additionally, these recommendations assume availability of time and resources to put towards service provision and research efforts. In reality, this may equate to resource reprioritization if there are not additional resources available. Therefore, competing public service priorities must be considered (eg. public transportation needs and infrastructure development). Finally, my recommendations focus on crime prevalence rather than crime severity. While homicide, for instance, is not one of the top ten most common crimes committed in Chicago, it is crucial that such severe crimes receive prioritization and attention. Additional data should be considered including crime severity, cost of services, and cost of corrective action.

Problem 4 – Question A – Of the types of crimes you have data for, which crime type is the most likely given the call came from 2111 S Michigan Avenue? What are the probabilities for each type of request?

Figure 22: Crimes Types and Probabilities on ‘S Michigan Ave’, 2017-2018

primary_type	number_of_cases	probability
THEFT	1490	0.241
BATTERY	1282	0.207
CRIMINAL DAMAGE	569	0.092
DECEPTIVE PRACTICE	539	0.087
OTHER OFFENSE	512	0.083
ASSAULT	410	0.066
ROBBERY	281	0.045
BURGLARY	212	0.034
MOTOR VEHICLE THEFT	207	0.033
NARCOTICS	150	0.024
CRIMINAL TRESPASS	148	0.024
WEAPONS VIOLATION	80	0.013
PROSTITUTION	76	0.012
OFFENSE INVOLVING CHILDREN	44	0.007
CRIM SEXUAL ASSAULT	40	0.006
PUBLIC PEACE VIOLATION	39	0.006
INTERFERENCE WITH PUBLIC OFFICER	34	0.005
SEX OFFENSE	19	0.003
HOMICIDE	17	0.003
LIQUOR LAW VIOLATION	13	0.002
STALKING	7	0.001
ARSON	7	0.001
OBSCENITY	5	0.001
KIDNAPPING	3	0
INTIMIDATION	2	0
HUMAN TRAFFICKING	2	0
GAMBLING	2	0
CONCEALED CARRY LICENSE VIOLATION	2	0

Of the types of crimes I have data for (2017-2018), the type of crime that is the most likely given the call came from ‘S Michigan Avenue’ is **theft at a 24.1% likelihood**. The probabilities for each type of request are displayed in the third column of this table (note: multiply by 100% to convert decimal to percent).

Problem 4 – Question B – Let’s now assume that a call comes in about Theft. Which is more likely – that the call came from Garfield Park or Uptown? How much more or less likely is it to be from Garfield Park vs. Uptown?

Figure 23: Theft in Uptown, E Garfield Park, and W Garfield Park, 2017-2018

	THEFT	probability
Uptown	1954	0.015097
East Garfield Park	1281	0.009897
West Garfield Park	1213	0.009372

If a call comes in about theft, it is more likely that the call came from Uptown than from East or West Garfield Park separately, but more likely that it came from **East or West Garfield Park (considered together) than from Uptown**. Considering all 129,427 calls made for theft between 2017 and 2018, it is **0.4% more likely** to have come from East or West Garfield Park (considered together) than from Uptown.

Problem 4 – Question C – If a call comes about Battery, how much more/less likely is it that the call came from Garfield Park versus Uptown?

Figure 24: Battery in Uptown and Garfield Park [Simulation]

	BATTERY	probability
Uptown	160	$160 / (160+100) = .6154$
Garfield Park	100	$100 / (160+100) = .3846$

Assuming that the 1000 calls mentioned in the prompt are historical and calls from Uptown and Garfield Park are routed to the same location, then a new call to this location about Battery is more likely to come from **Uptown** than from Garfield Park, by **23.1%** (which is equal to $(.6154 - .3846) * 100\%$). Given that the call is definitely about Battery, then only the historical calls about Battery are of interest to these calculations rather than all previous calls regardless of crime type.