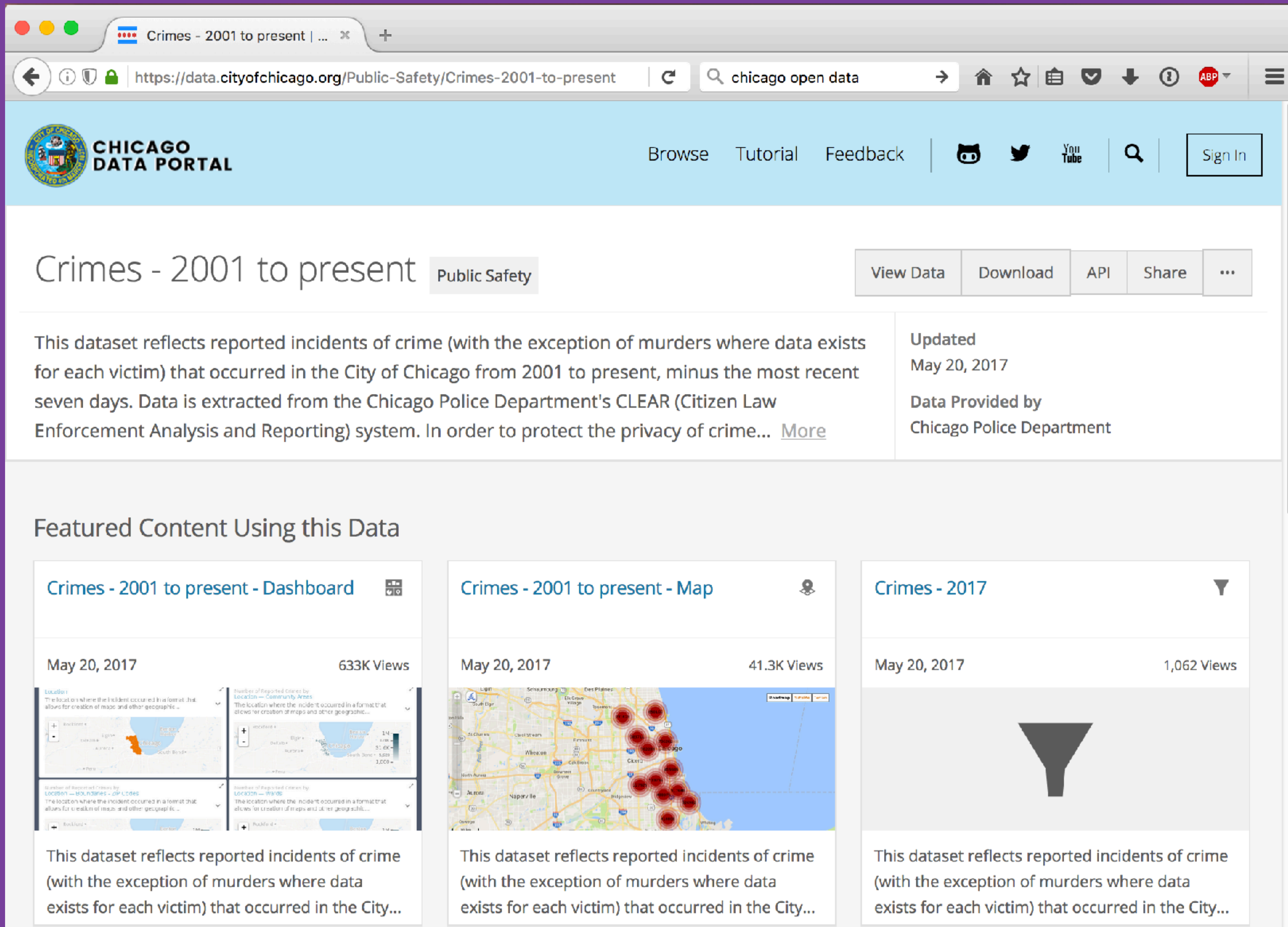
Crime in Chicago

Northwestern University
MSPA Predict 498-55

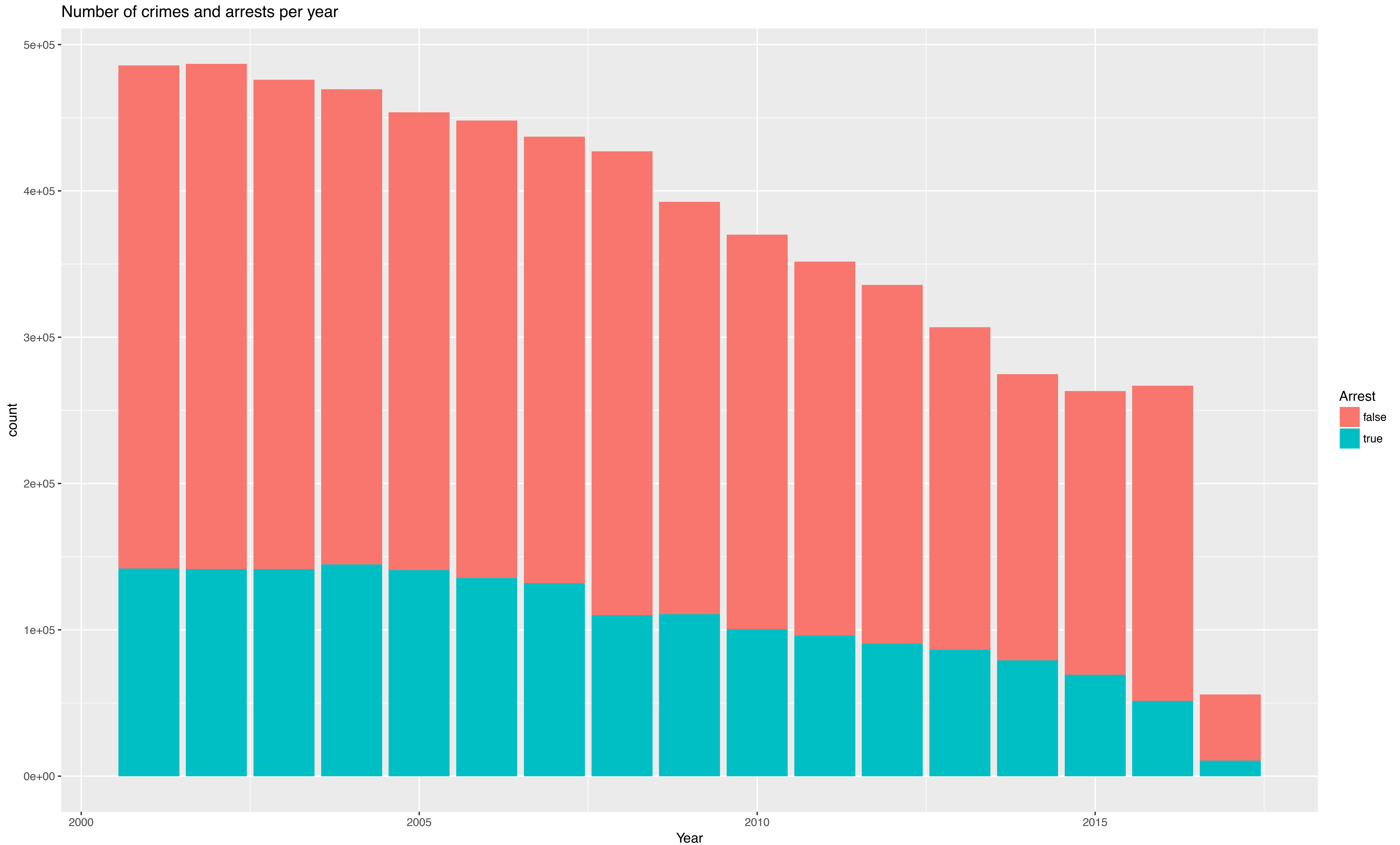
May 28, 2017

Kent Merdes
Maryna Obreski
Russ Conte
Shashikala Kadanapuradoddi Lingaiah

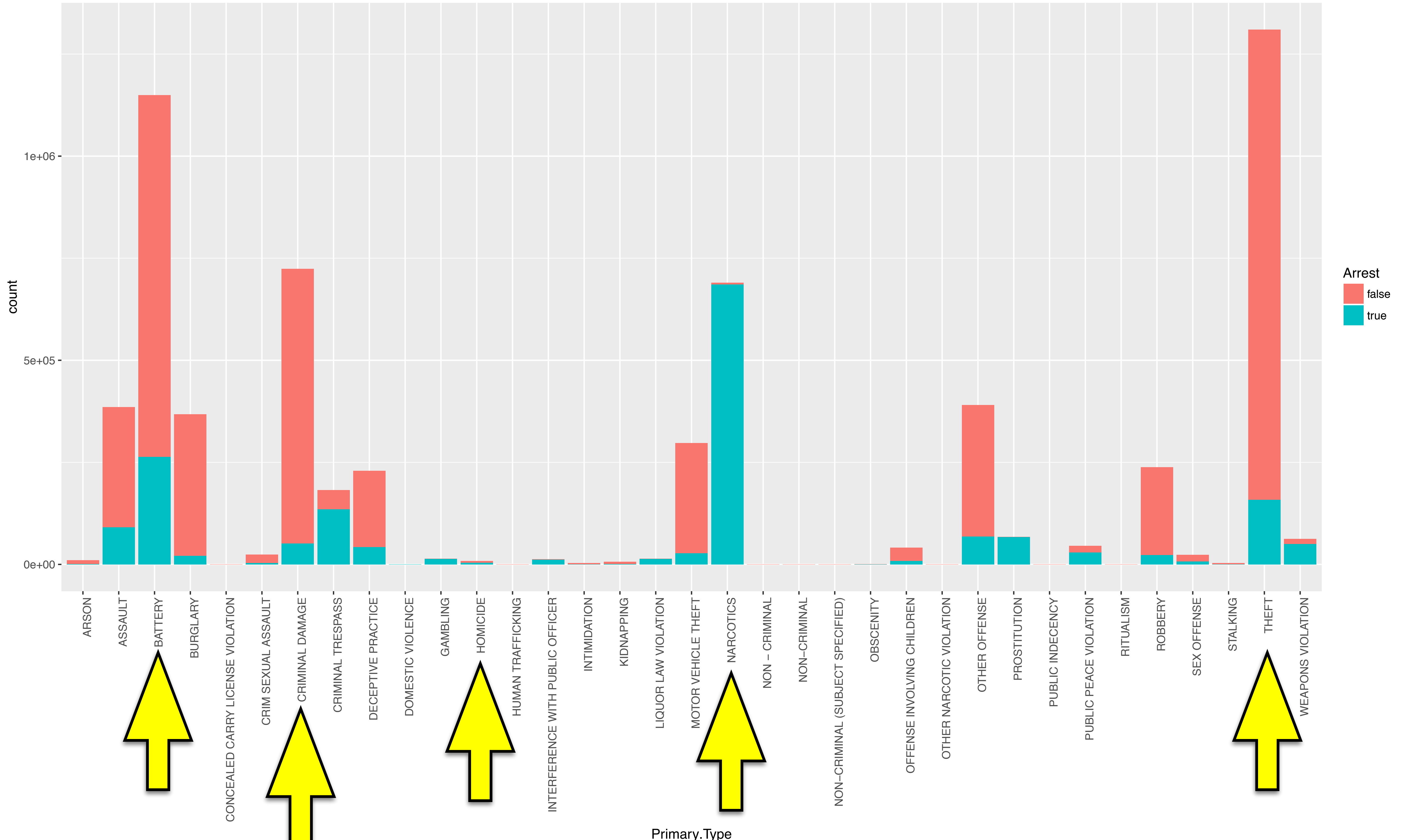
The Chicago Open Data Portal: Crimes – January 1, 2001–Present



Total number of crimes in Chicago by year, January 1, 2001–May 14, 2017

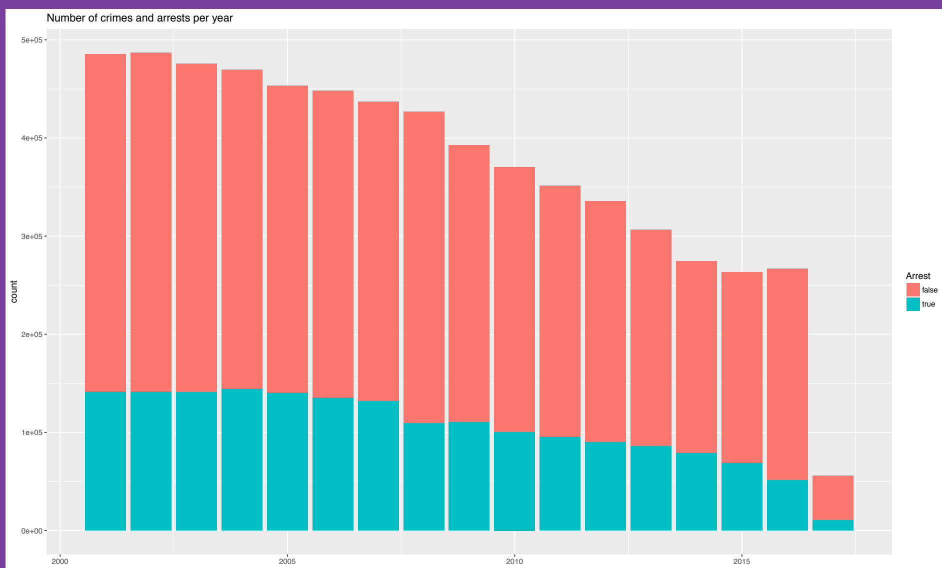


Total number of crimes by Primary Type and arrest

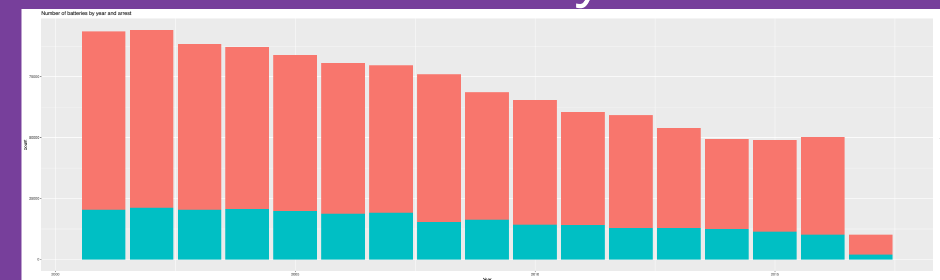


Total number of crimes by year, type of crime, arrest

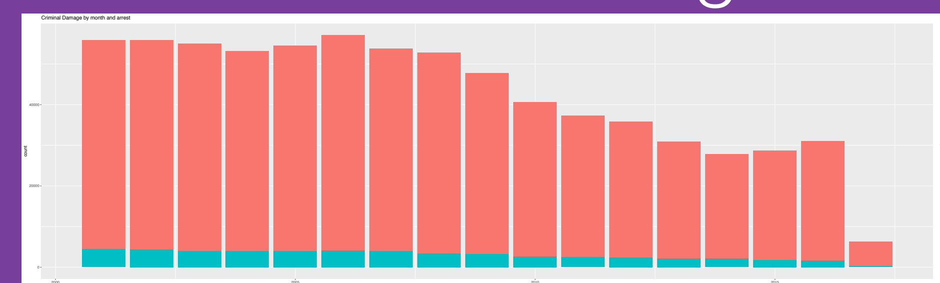
All crimes



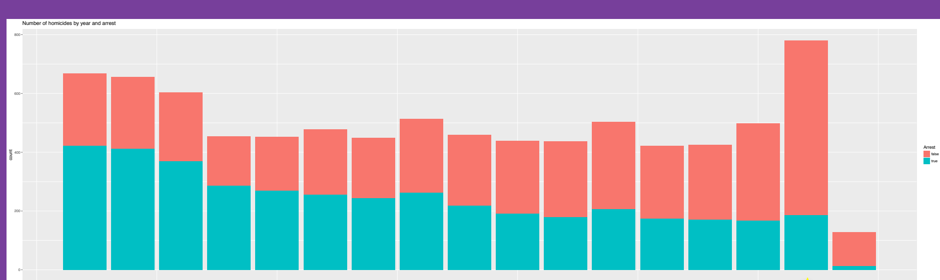
Battery



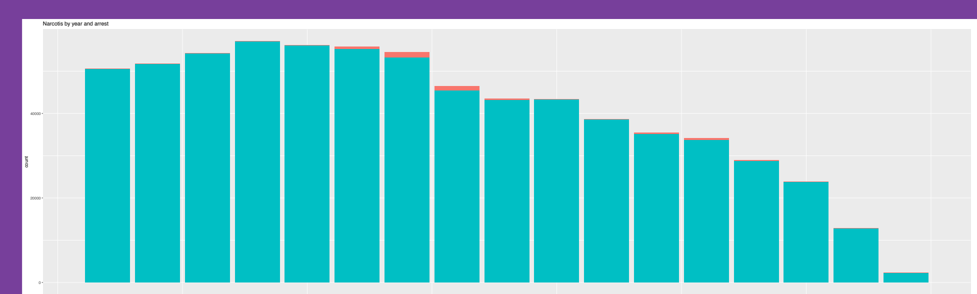
Criminal Damage



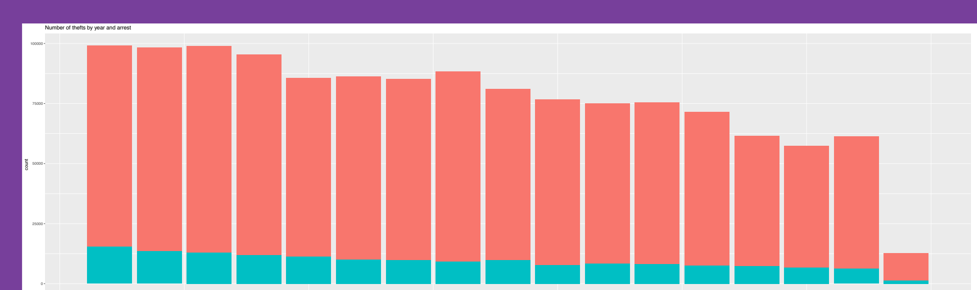
Homicides



Narcotics

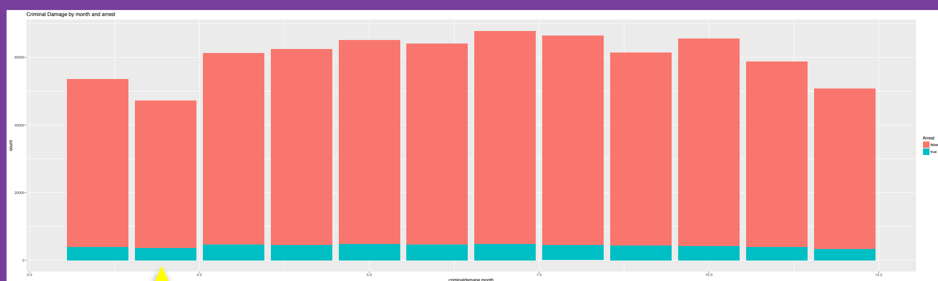


Theft

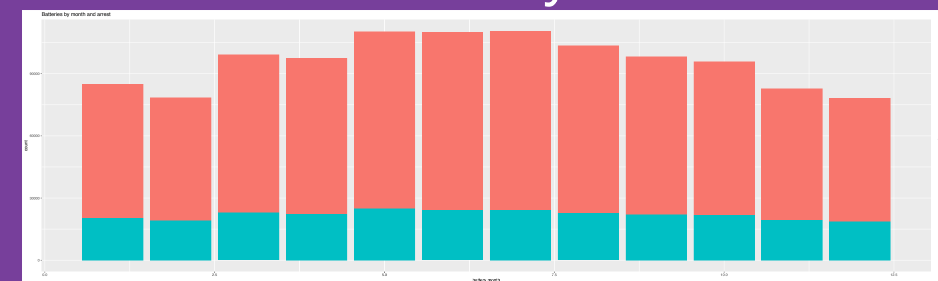


Total number of crimes by month, type of crime, arrest

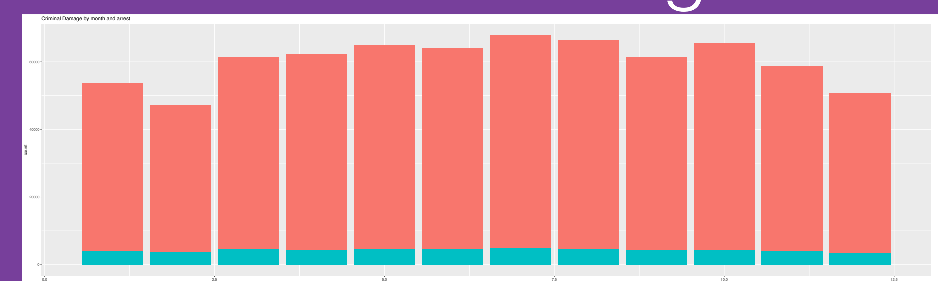
All crimes



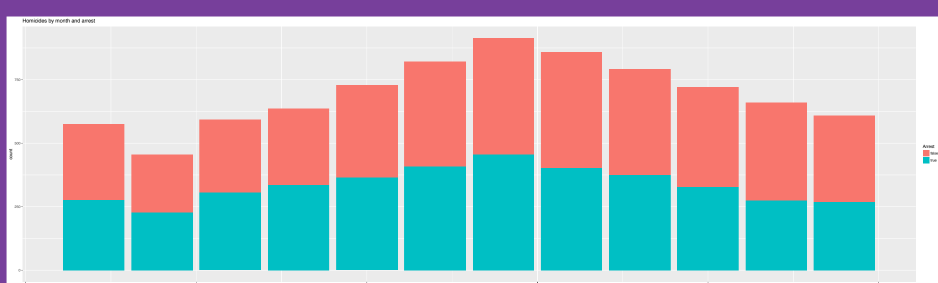
Battery



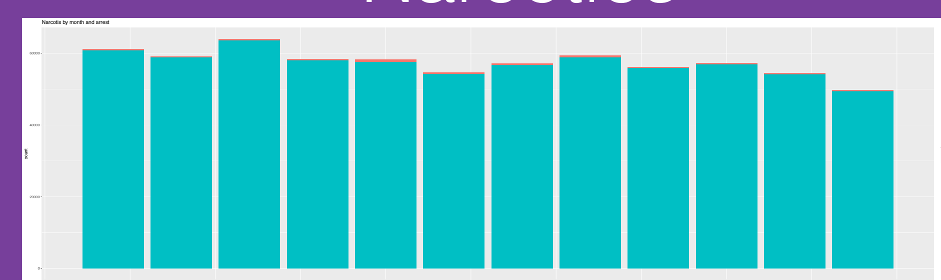
Criminal Damage



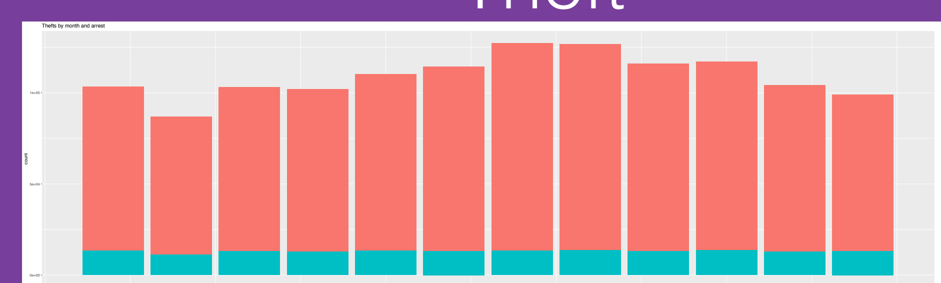
Homicides



Narcotics

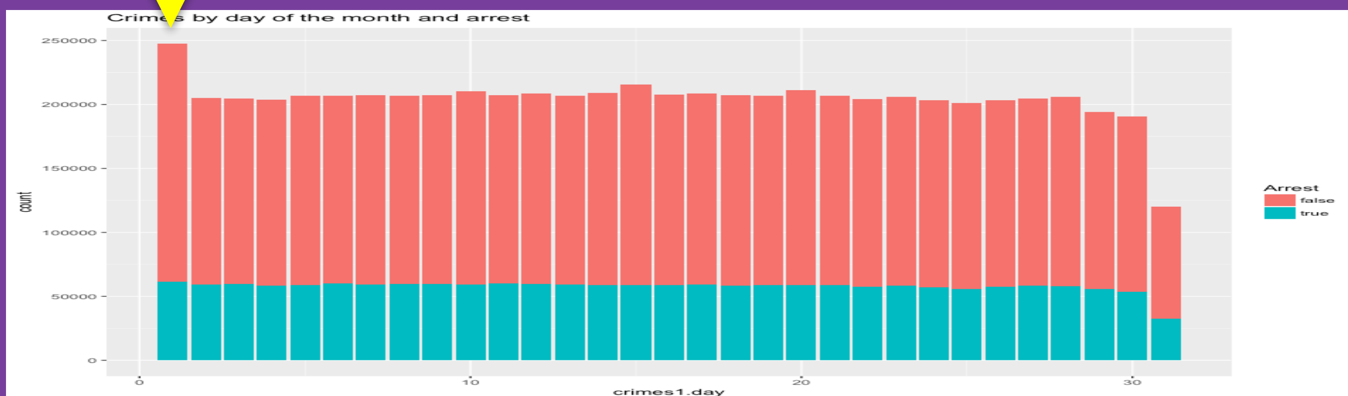


Theft

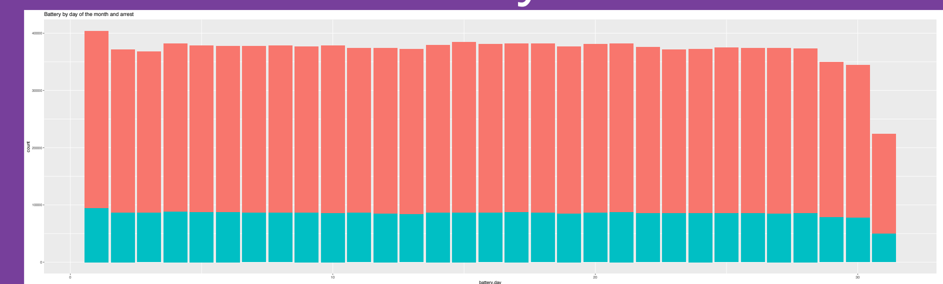


Total number of crimes by day of the month, type of crime, arrest

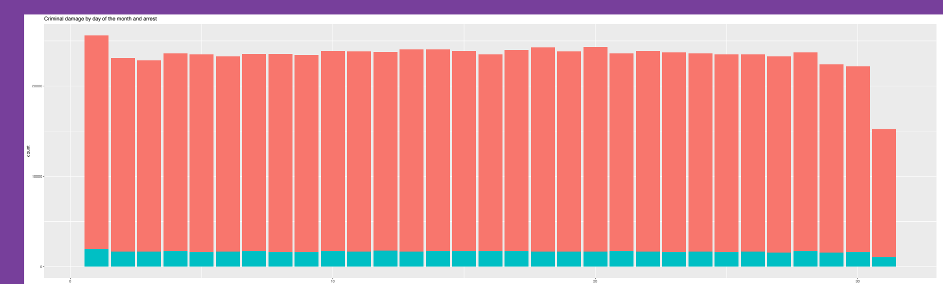
All crimes



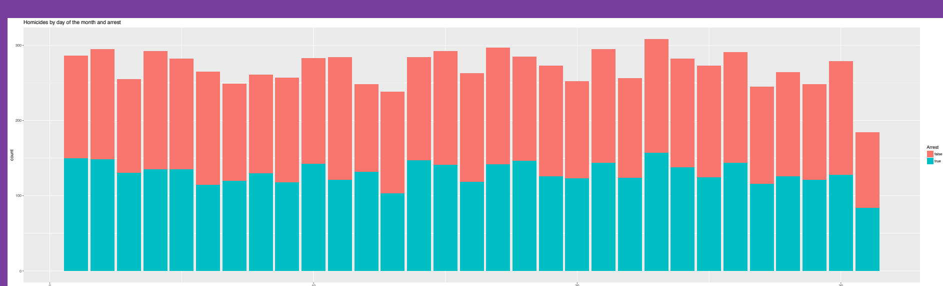
Battery



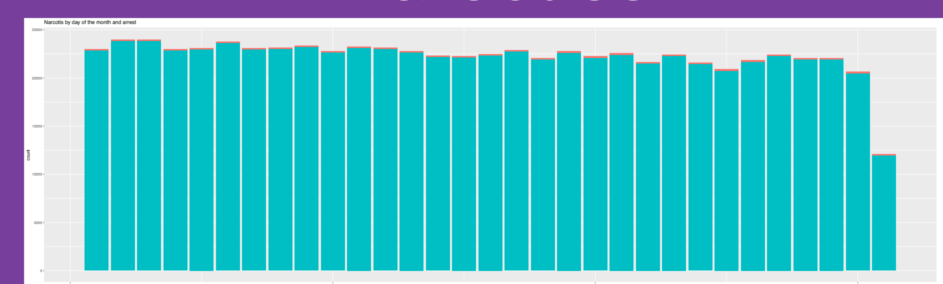
Criminal Damage



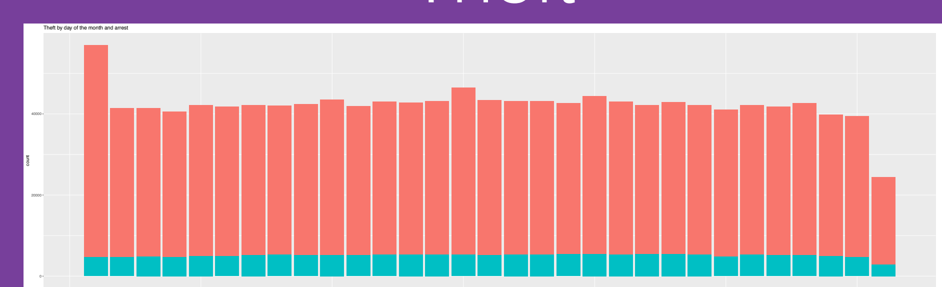
Homicides



Narcotics

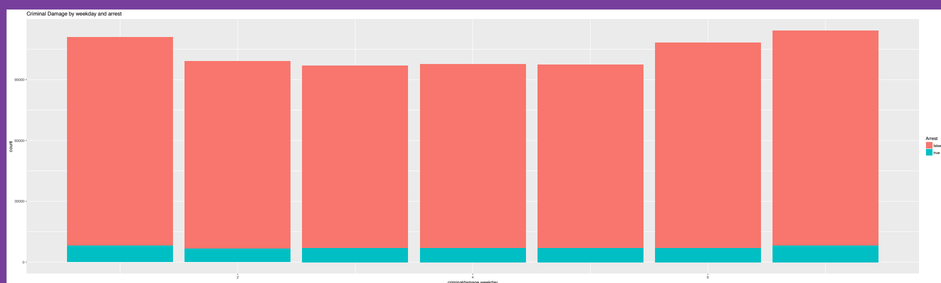


Theft

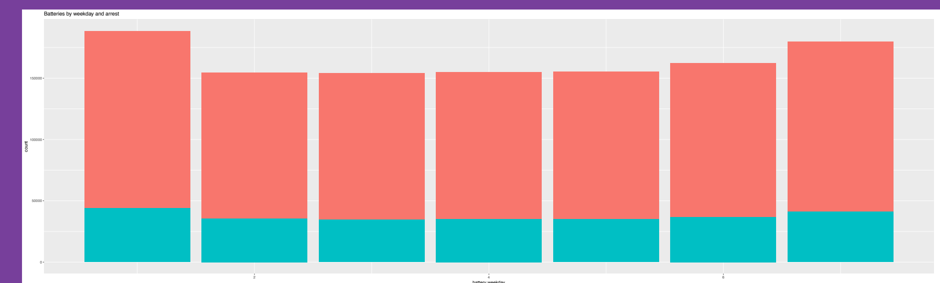


Total number of crimes by weekday, type of crime, arrest

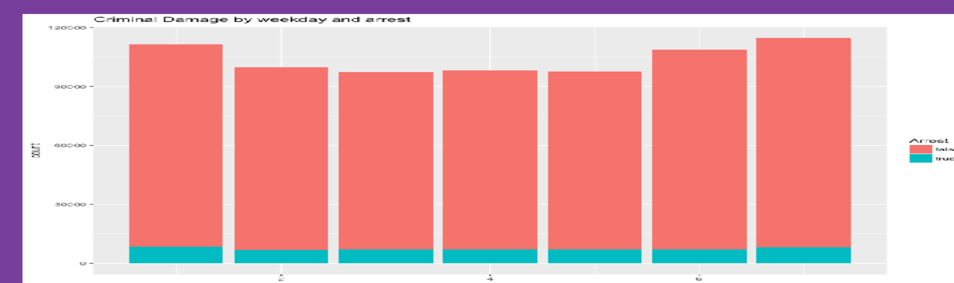
All crimes



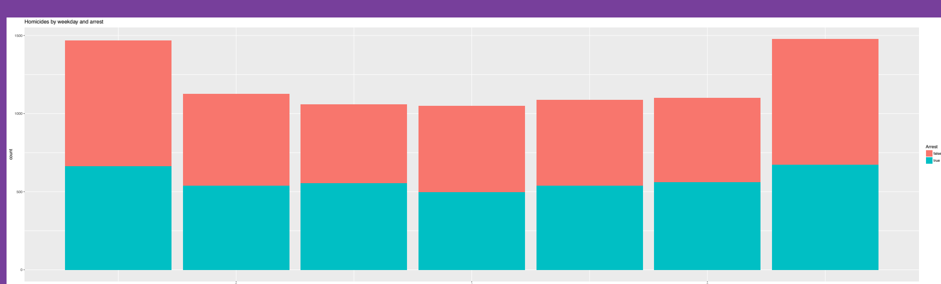
Battery



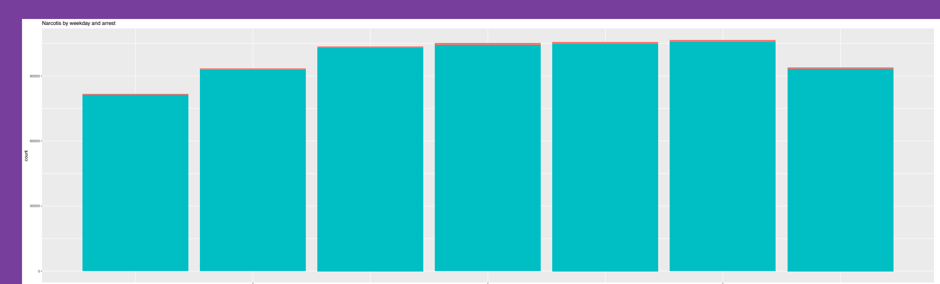
Criminal Damage



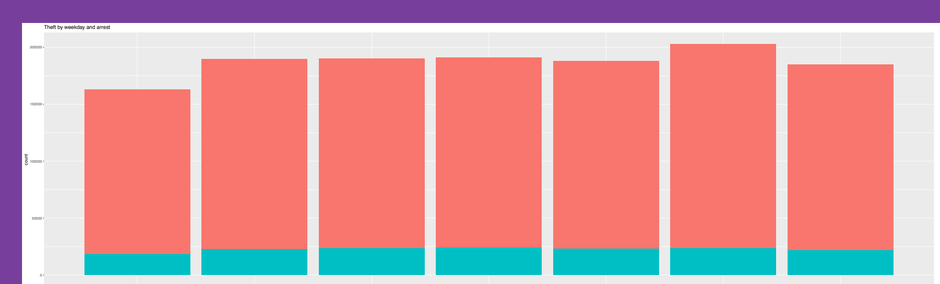
Homicides



Narcotics

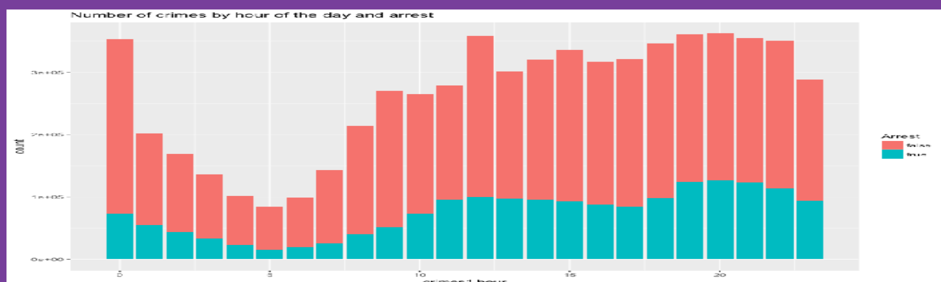


Theft

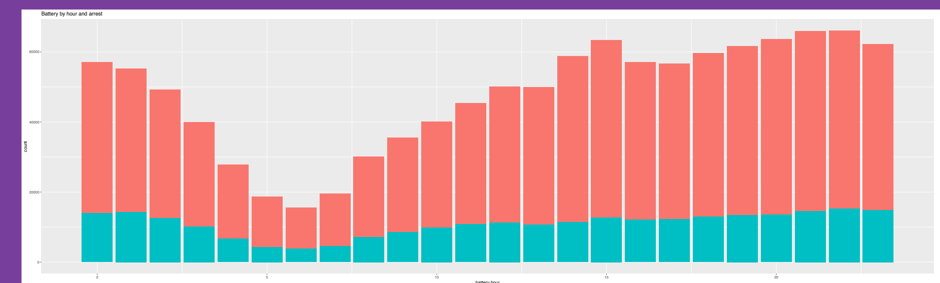


Total number of crimes by hour, type of crime, arrest

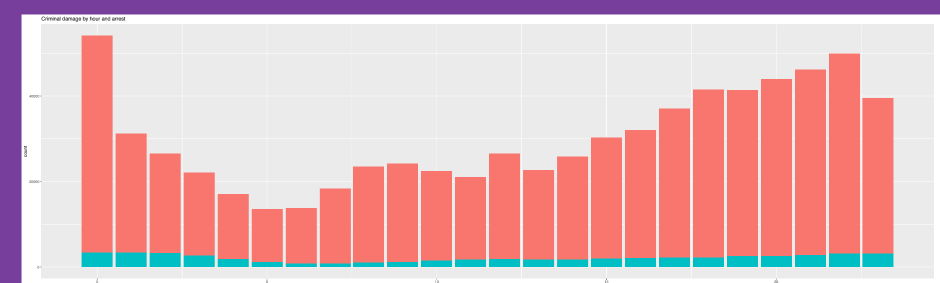
All crimes



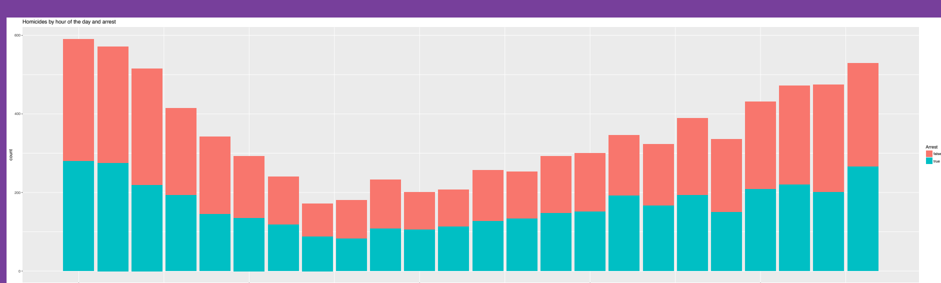
Battery



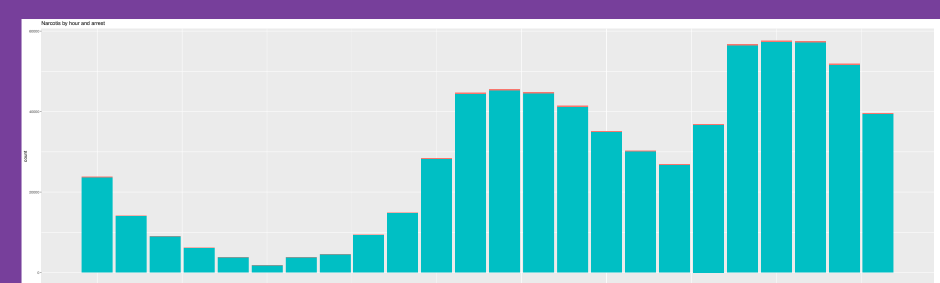
Criminal Damage



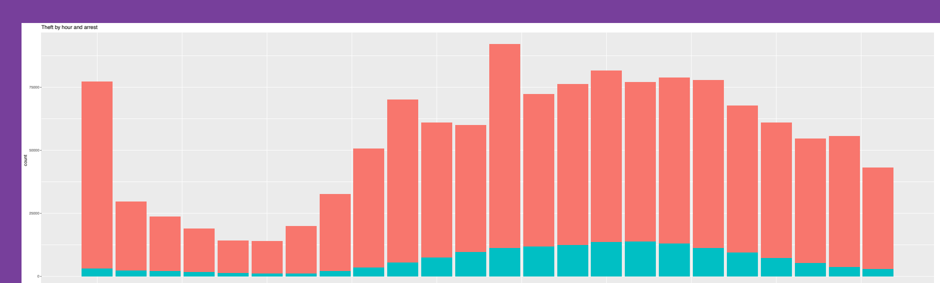
Homicides



Narcotics



Theft



Chicago Crimes By Primary Type and By Year (Jan 2001–Apr 2017)

	Year of Date																
Primary Type (group)	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
THEFT	99,264	98,327	98,875	95,463	85,685	86,238	85,156	88,430	80,972	76,749	75,142	75,456	71,522	61,536	57,306	61,436	17,809
BATTERY	93,447	94,152	88,376	87,134	83,964	80,665	79,591	75,923	68,462	65,400	60,458	59,133	54,003	49,446	48,908	50,263	14,481
CRIMINAL DAMAGE	55,851	55,940	55,011	53,164	54,548	57,124	53,749	52,841	47,723	40,653	37,332	35,854	30,853	27,798	28,671	31,001	8,851
NARCOTICS	50,567	51,789	54,288	57,060	56,234	55,813	54,454	46,507	43,543	43,393	38,606	35,487	34,129	28,974	23,837	12,989	3,123
OTHER OFFENSE	29,656	32,599	31,147	29,531	28,028	27,100	26,863	26,533	25,601	22,012	20,188	17,478	17,987	16,969	17,537	17,153	5,720
ASSAULT	31,384	31,521	29,477	28,850	27,066	25,945	26,314	25,447	22,860	21,535	20,411	19,898	17,971	16,896	17,040	18,725	5,596
BURGLARY	26,011	25,623	25,156	24,564	25,504	24,324	24,858	26,218	26,766	26,422	26,619	22,843	17,894	14,569	13,182	14,278	4,017
MOTOR VEHICLE THEFT	27,549	25,121	22,748	22,805	22,497	21,818	18,573	18,881	15,482	19,028	19,387	16,492	12,581	9,914	10,073	11,318	3,578
ROBBERY	18,441	18,522	17,332	15,978	16,047	15,968	15,450	16,703	15,979	14,273	13,982	13,485	11,820	9,800	9,639	11,958	3,411
DECEPTIVE PRACTICE	14,892	13,710	13,444	13,234	13,551	13,557	14,119	14,858	13,763	12,440	12,534	13,469	13,510	15,370	15,518	17,894	5,233
CRIMINAL TRESPASS	13,240	13,880	14,807	15,913	16,655	14,505	13,699	12,310	10,851	9,401	8,659	8,215	8,135	7,538	6,401	6,307	2,137
PROSTITUTION	6,026	6,408	6,214	7,476	6,124	7,034	6,087	5,141	3,940	2,485	2,424	2,204	1,652	1,625	1,322	800	281
WEAPONS VIOLATION	4,274	4,281	4,211	4,297	4,106	3,821	3,554	3,877	4,158	3,704	3,880	3,904	3,246	3,114	3,362	3,443	1,324
PUBLIC PEACE VIOLATION	2,750	2,457	2,430	2,495	2,730	3,068	3,315	3,013	3,147	3,538	3,095	3,007	3,135	2,903	2,421	1,604	463
OFFENSE INVOLVING CHILD..	2,237	2,571	3,031	3,072	2,875	2,744	2,862	2,606	2,548	2,521	2,319	2,186	2,318	2,344	2,210	2,249	629
CRIM SEXUAL ASSAULT	1,795	1,829	1,588	1,568	1,542	1,459	1,532	1,517	1,409	1,345	1,465	1,407	1,267	1,310	1,348	1,469	434
SEX OFFENSE	2,229	2,167	2,068	1,802	1,803	1,561	1,521	1,482	1,244	1,109	1,069	1,050	1,015	953	952	932	239
GAMBLING	934	971	1,088	1,122	1,078	1,368	1,409	1,199	991	927	736	724	596	393	310	189	32
LIQUOR LAW VIOLATION	1,637	1,414	1,311	985	1,005	1,135	1,170	912	746	736	619	573	465	397	292	227	64
INTERFERENCE WITH PUBLI..	406	361	408	531	615	758	677	580	573	796	1,048	1,228	1,281	1,398	1,308	934	331
ARSON	1,010	1,032	955	778	691	726	712	644	616	522	504	469	364	397	451	518	146
HOMICIDE	644	637	570	443	445	459	430	488	441	410	422	479	409	417	471	744	167
KIDNAPPING	933	829	705	481	389	339	330	359	293	313	266	236	243	220	190	201	54
INTIMIDATION	279	337	364	349	258	276	255	261	231	197	171	156	134	117	121	132	47
STALKING	203	200	247	215	192	186	213	190	167	189	181	207	153	140	154	173	68
OBSCENITY	19	26	16	13	19	17	11	13	21	33	40	26	24	37	46	53	19
NON-CRIMINAL			1						1	1		8	7	28	35	54	9
PUBLIC INDECENCY	9	8	6	9	4	4	5	4	10	7	13	17	10	10	14	10	3

Chicago Crimes By Year/Month and By Location (Jan 2001–Apr 2017)

Primary Type

(All)

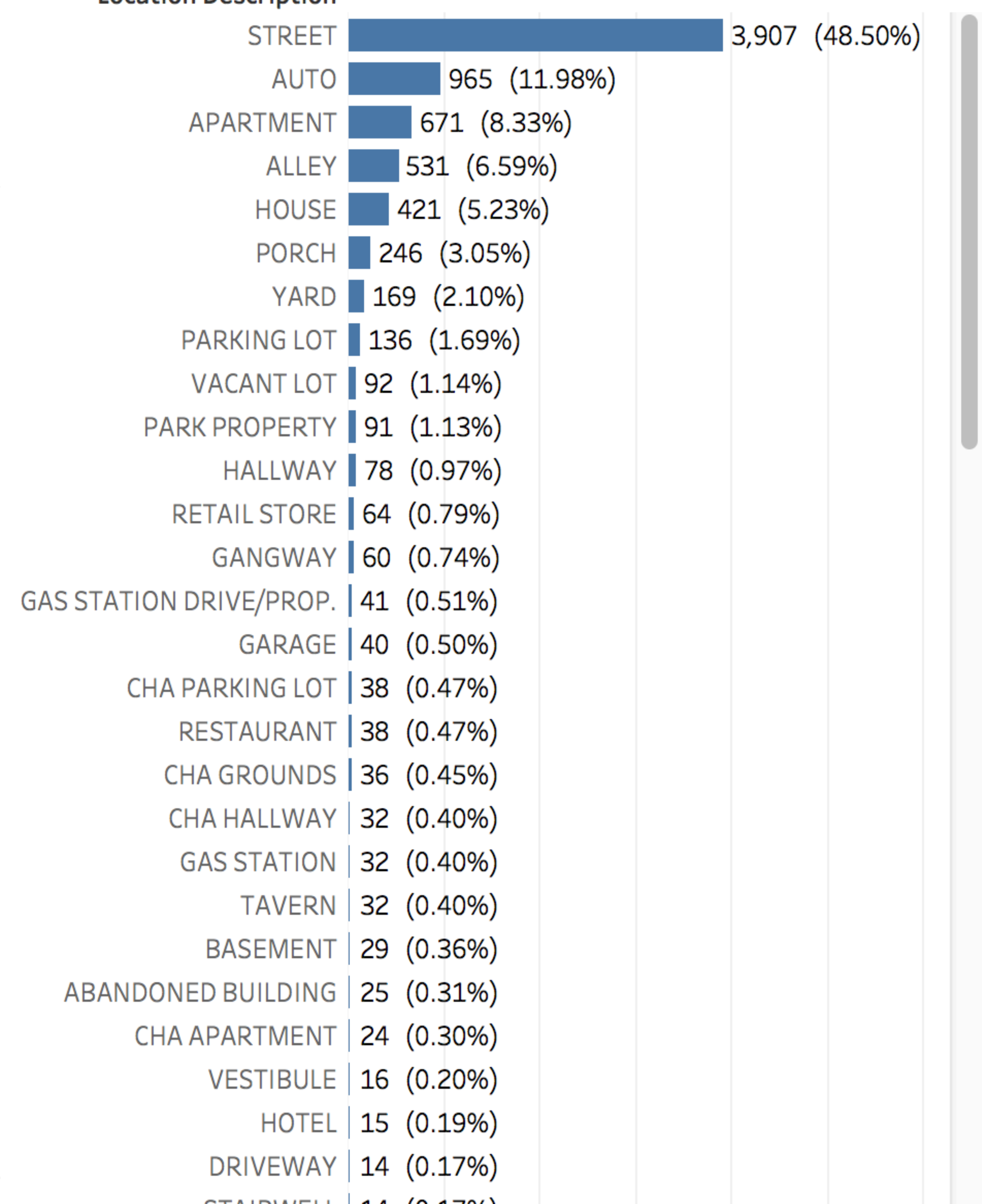
Date

1/1/01 10:40:00 AM — 4/12/17 1:04:00 PM

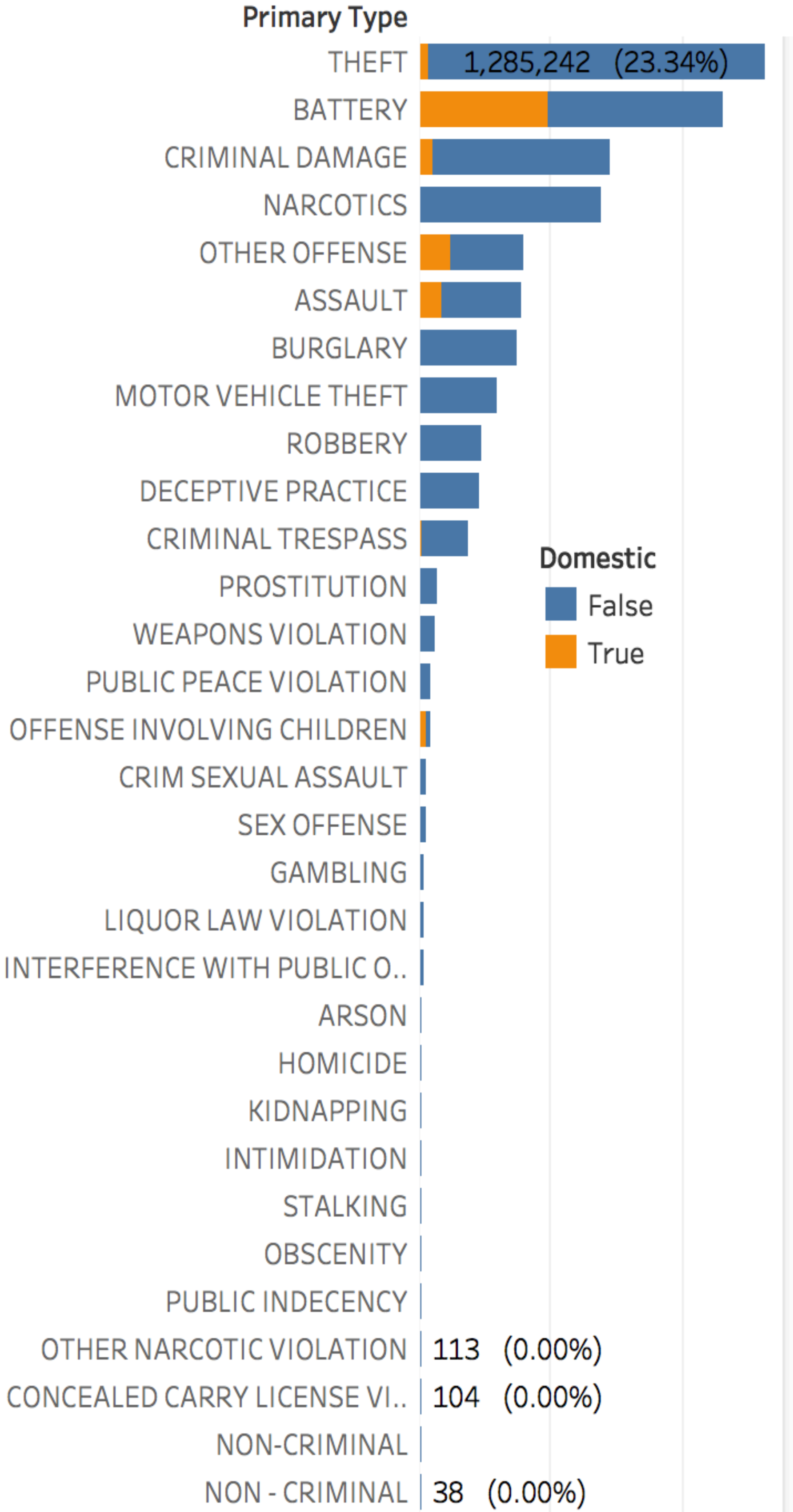
Date

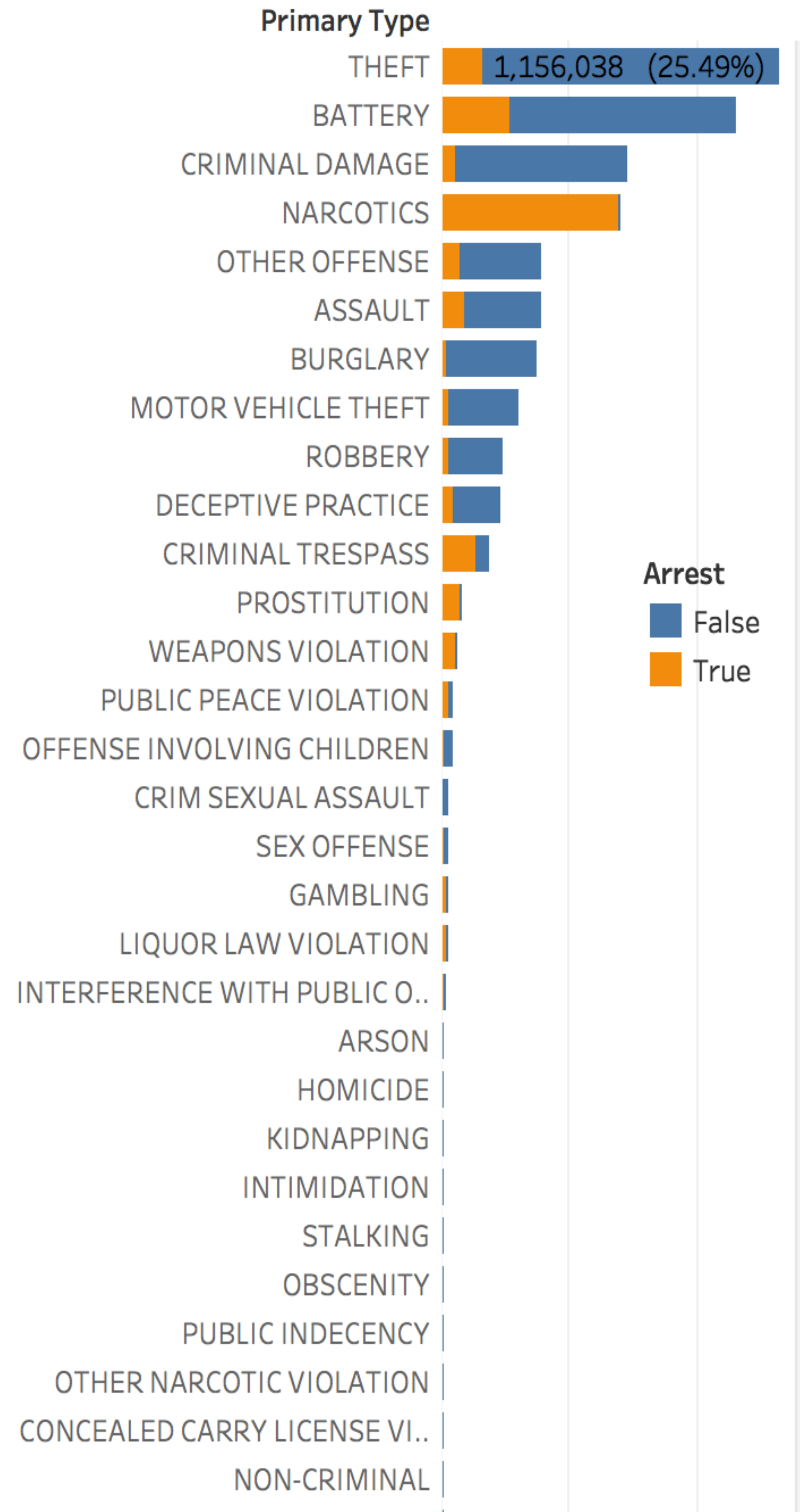
Year of Date	January	February	March	April	May	June	July	August	September	October	November	December
2001	38,089	33,777	40,552	40,077	41,819	41,720	44,686	44,028	41,499	43,020	39,593	36,843
2002	38,397	33,906	38,580	40,034	42,910	42,832	46,010	44,210	42,381	43,143	37,150	37,169
2003	36,697	32,506	38,648	39,784	41,187	40,810	43,408	44,257	41,421	43,325	38,049	35,786
2004	35,095	34,063	39,851	38,266	40,778	41,130	43,233	43,043	41,210	41,520	36,835	34,318
2005	33,841	31,984	36,903	38,869	40,464	40,047	41,802	41,539	39,622	39,958	35,974	32,663
2006	36,737	31,294	36,972	36,473	39,609	38,807	41,536	40,491	37,765	38,651	35,563	34,133
2007	34,066	27,133	36,341	35,627	40,083	38,981	40,985	39,843	38,253	39,587	34,418	31,606
2008	33,343	29,037	33,970	35,581	38,058	37,689	40,478	40,531	37,435	37,872	33,550	29,405
2009	30,200	28,219	33,676	32,555	35,228	34,257	35,662	35,809	33,857	33,521	31,382	28,180
2010	29,162	24,934	32,328	31,654	33,386	32,694	33,499	34,135	31,870	32,385	28,871	25,228
2011	27,145	22,217	28,682	29,087	31,566	32,291	33,239	32,570	29,908	30,256	27,612	27,002
2012	26,207	23,850	28,535	27,167	30,064	31,048	31,947	30,012	27,730	27,938	26,013	25,186
2013	25,369	21,377	24,924	25,497	27,958	27,328	28,593	28,619	26,297	25,427	23,522	21,820
2014	19,882	17,986	22,130	22,846	24,809	25,358	26,485	25,804	23,820	23,917	20,693	20,910
2015	20,684	16,306	21,574	21,623	23,581	23,069	24,109	24,696	23,006	22,994	20,508	21,023
2016	20,434	18,616	21,900	20,991	23,378	23,880	24,705	24,737	23,486	23,553	21,470	19,955
2017	21,280	18,716	19,827	18,467								

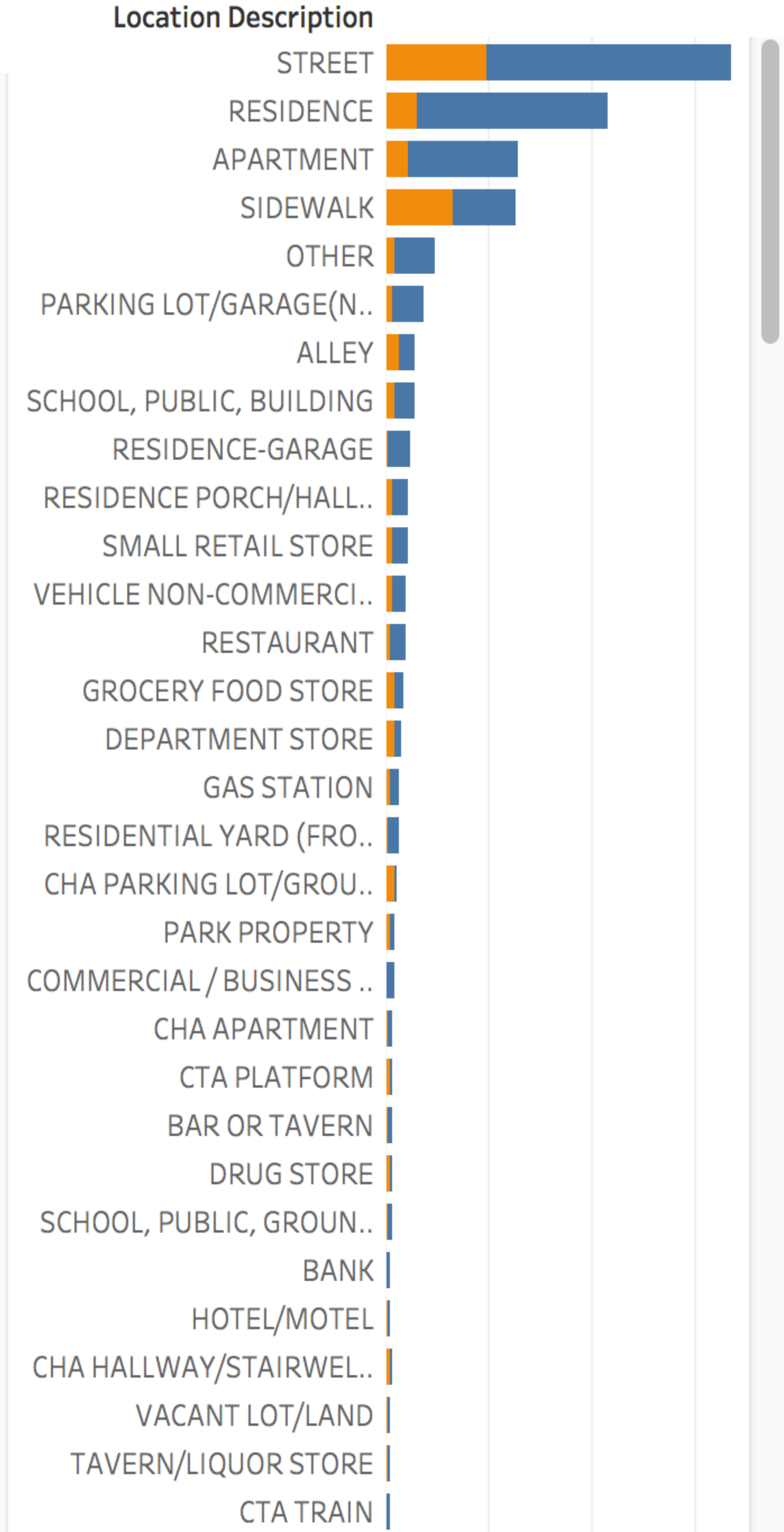
Location Description



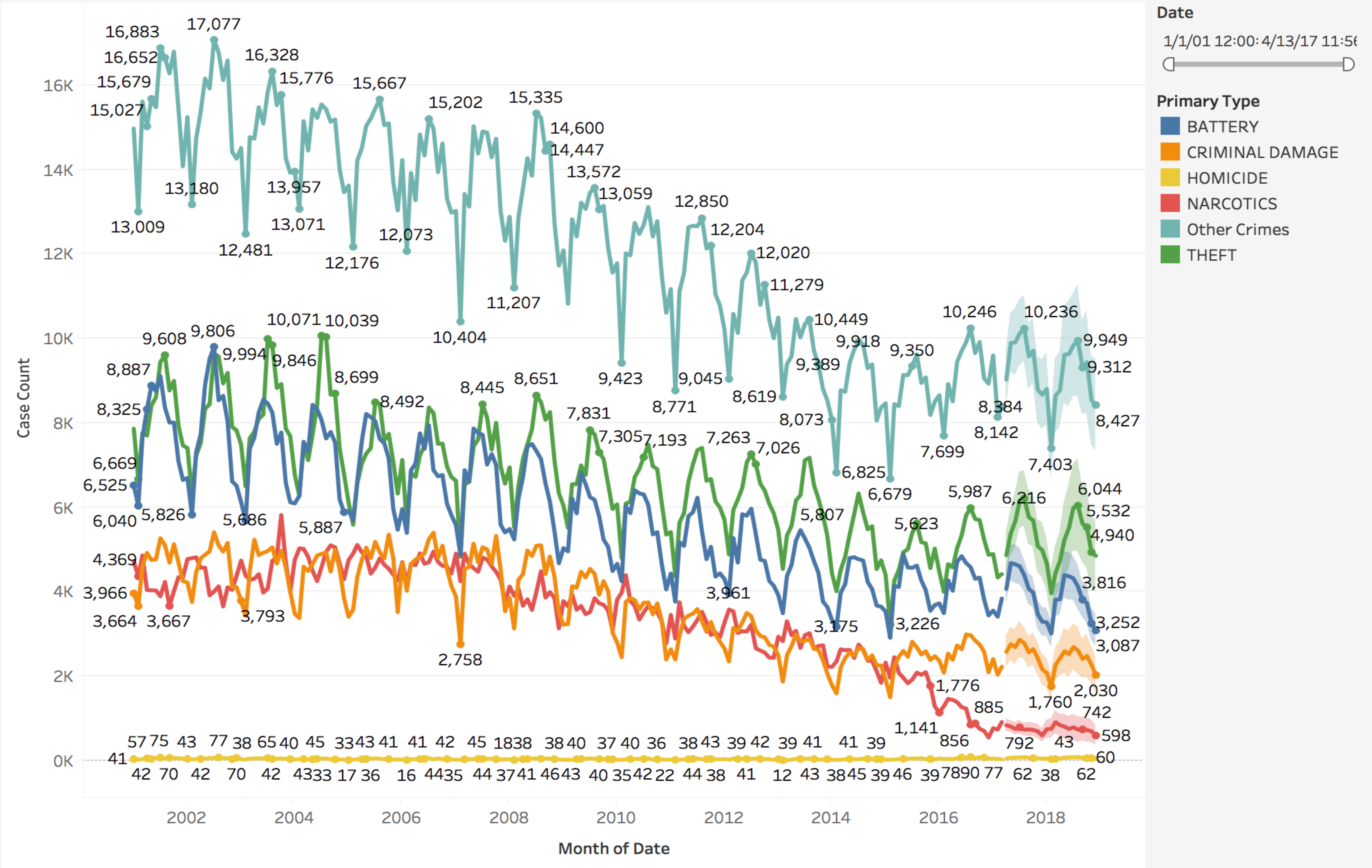
Chicago Crimes By Primary Type and By Location (Jan 2001–Apr 2017)







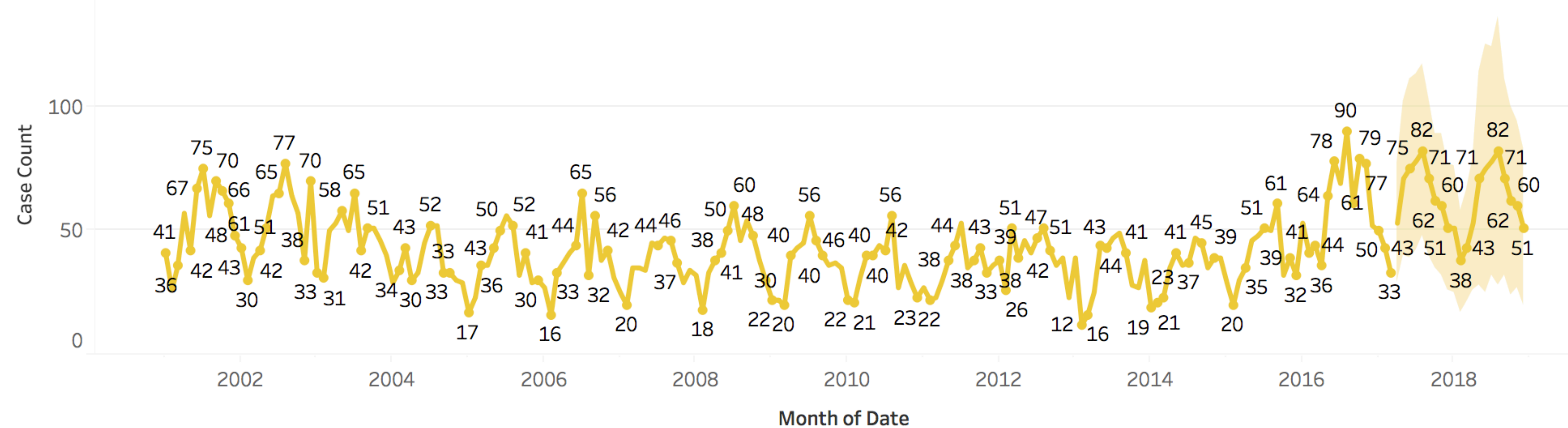
Chicago Crimes Forecasting (Jan 2001–Apr 2017)

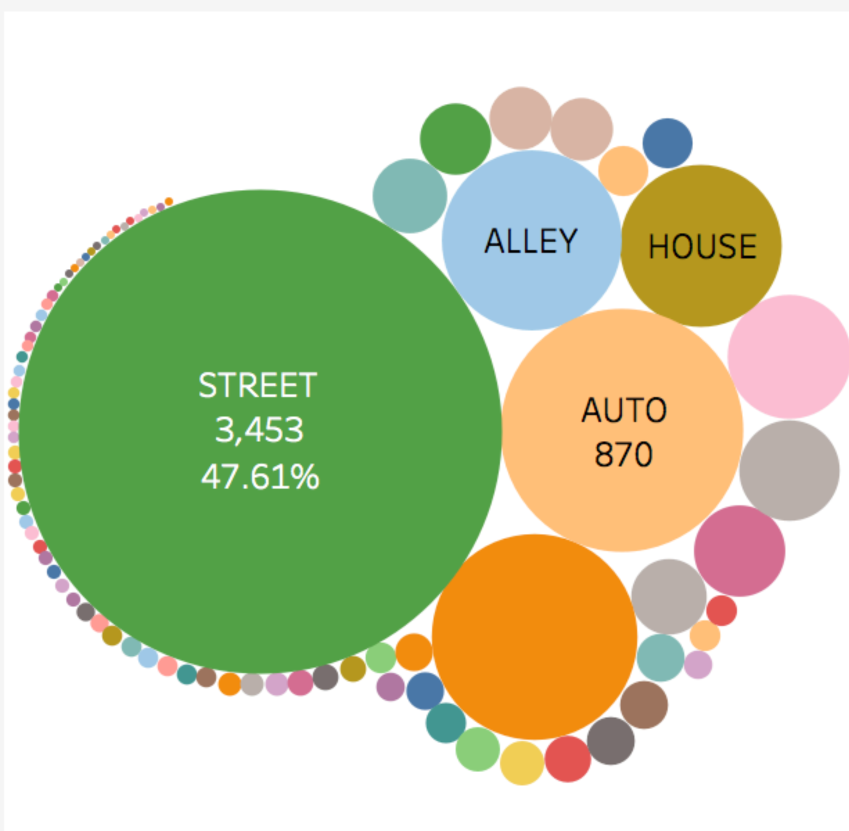


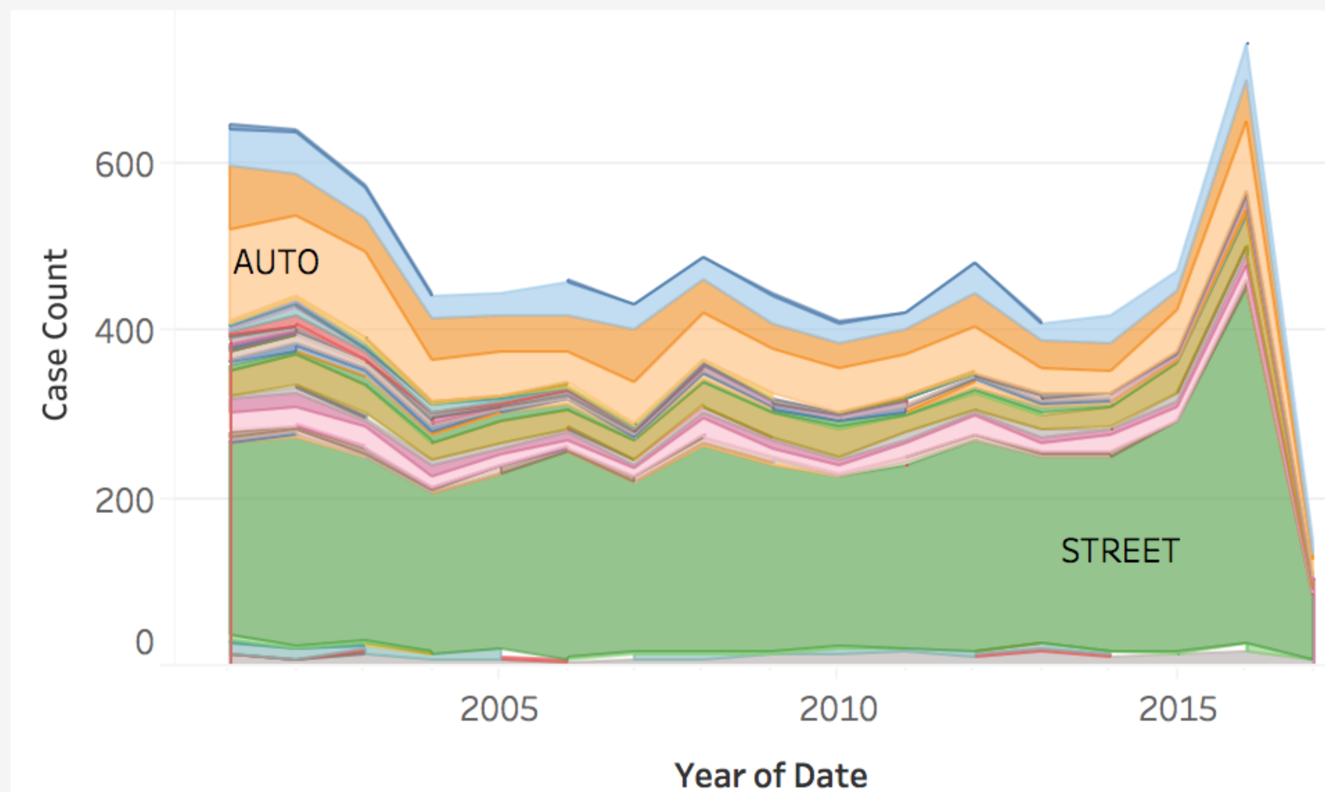
Crime in Chicago | **Data Visualizations** | Time Series Models | Machine Learning | Wrap-Up

Chicago Homicides Forecasting and By Location (Jan 2001–Apr 2017)

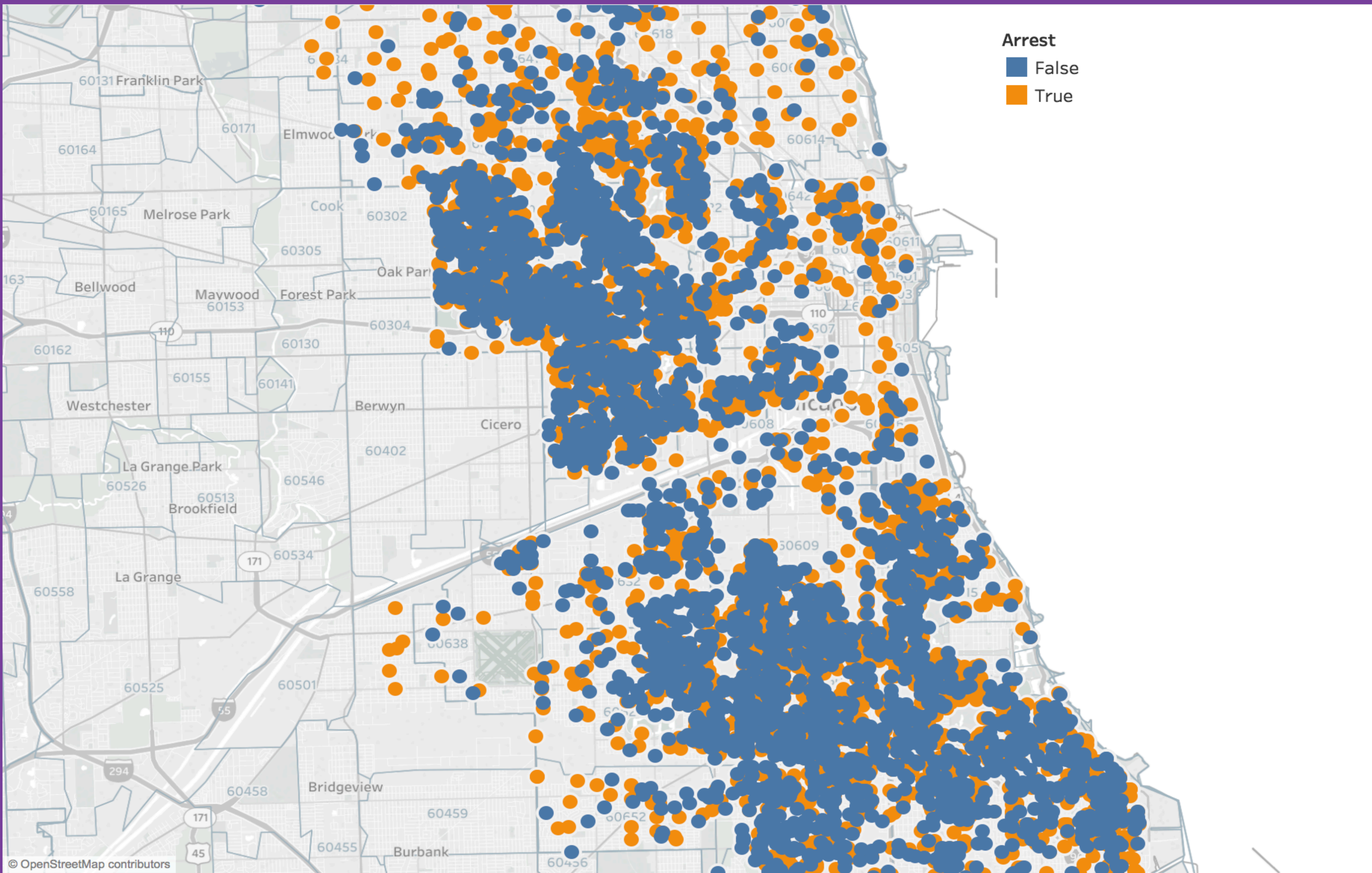
Primary Ty.. HOMICIDE **Date** 1/1/01 12:00:00 AM 2/23/16 11:59:59 PM



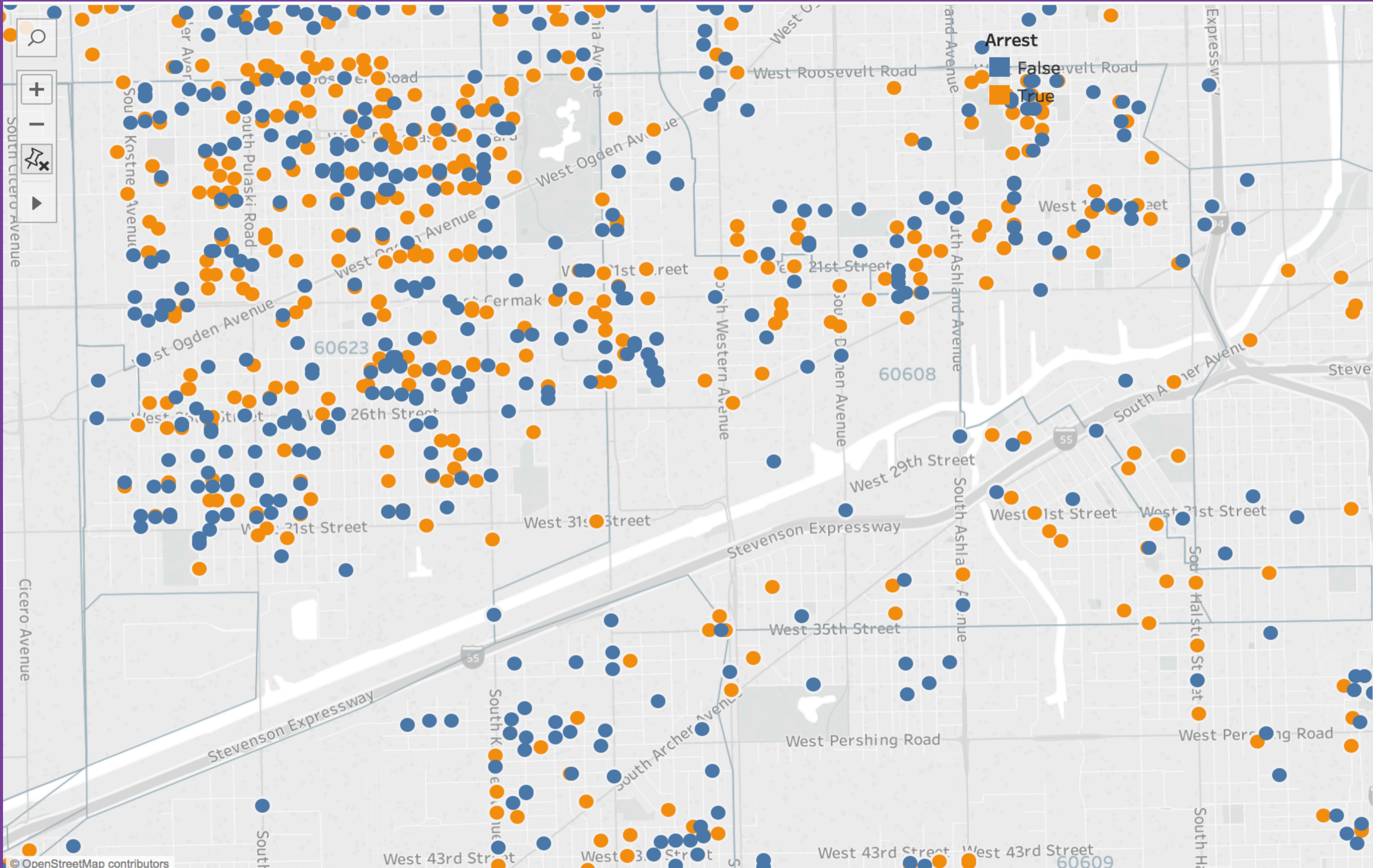




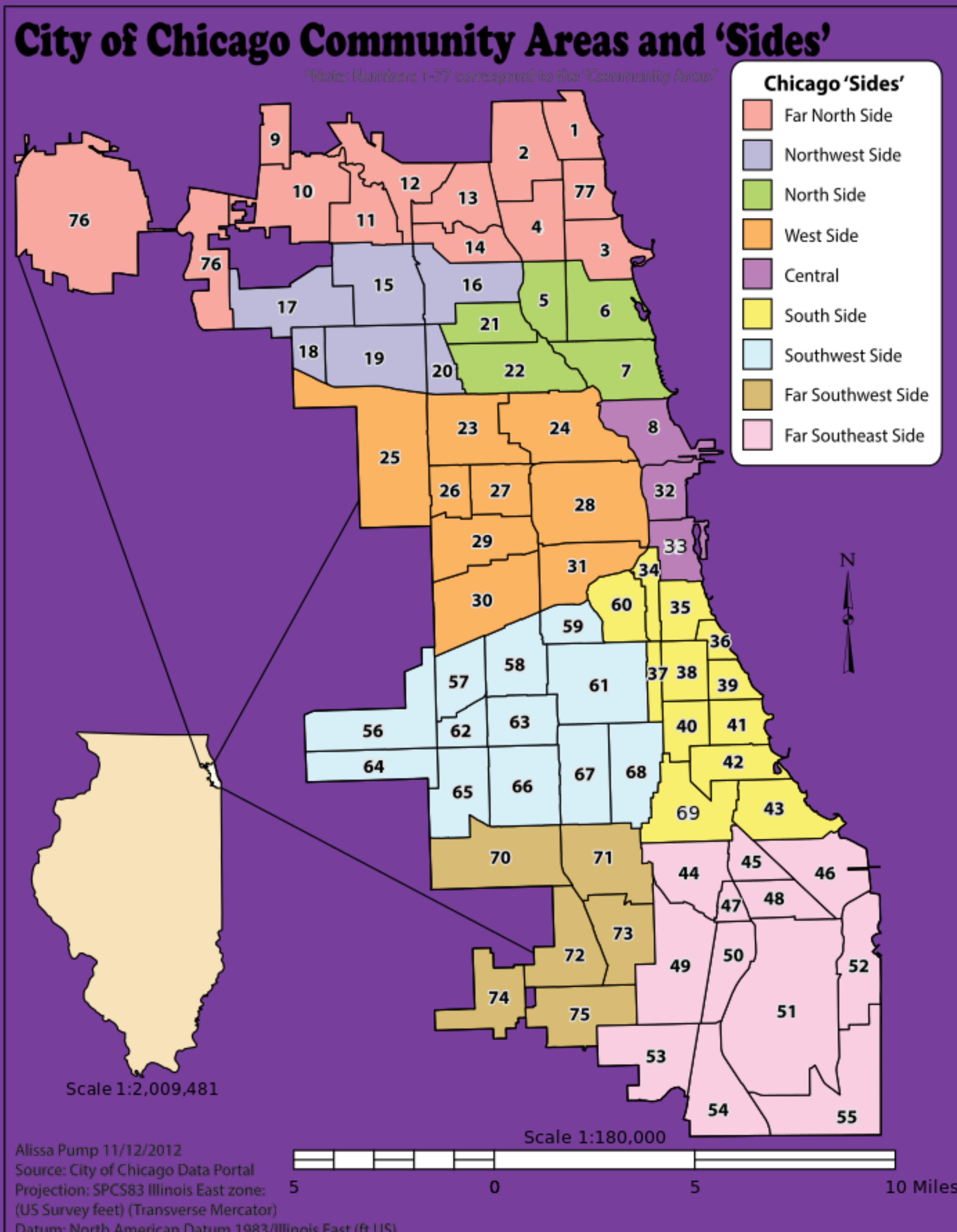
Chicago Homicides By Location (Jan 2001–Apr 2017)

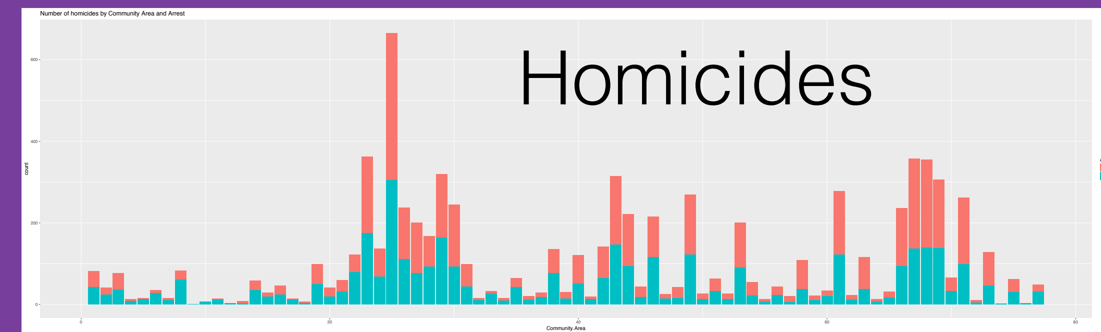


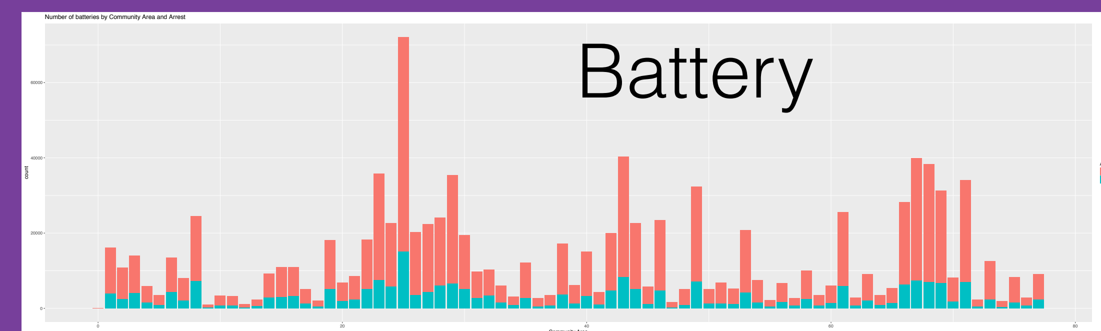
Chicago Homicides By Location (Jan 2001–Apr 2017)

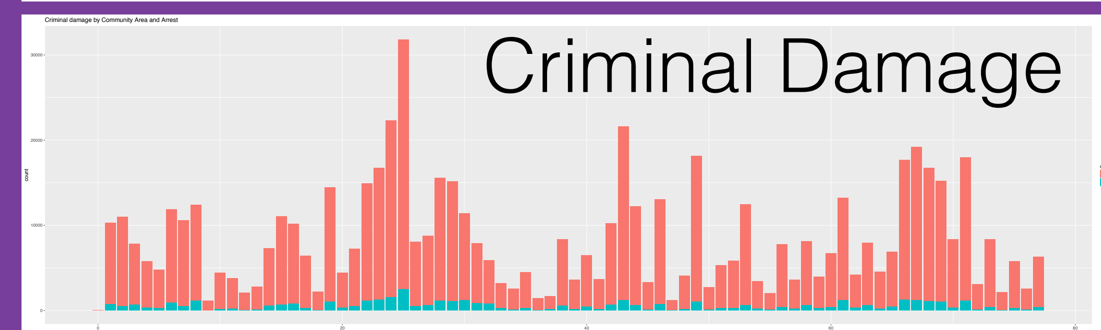


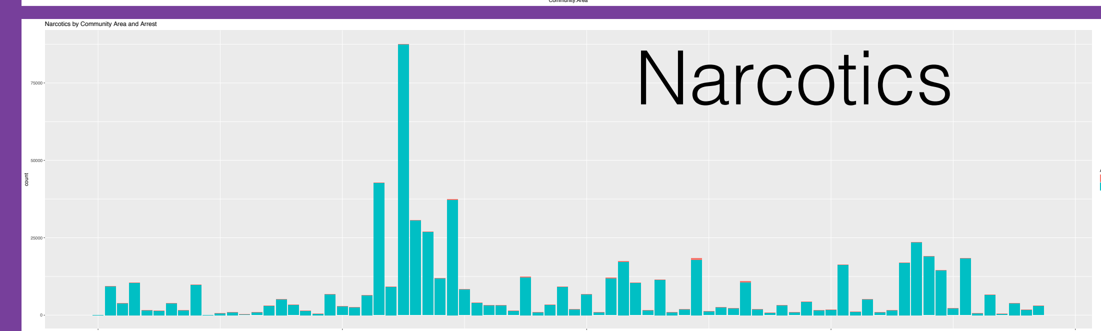
Total number of crimes by community area, type of crime, arrest

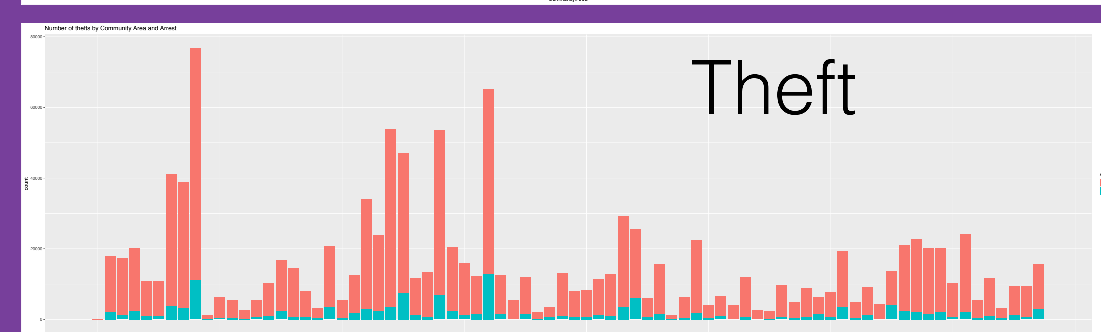




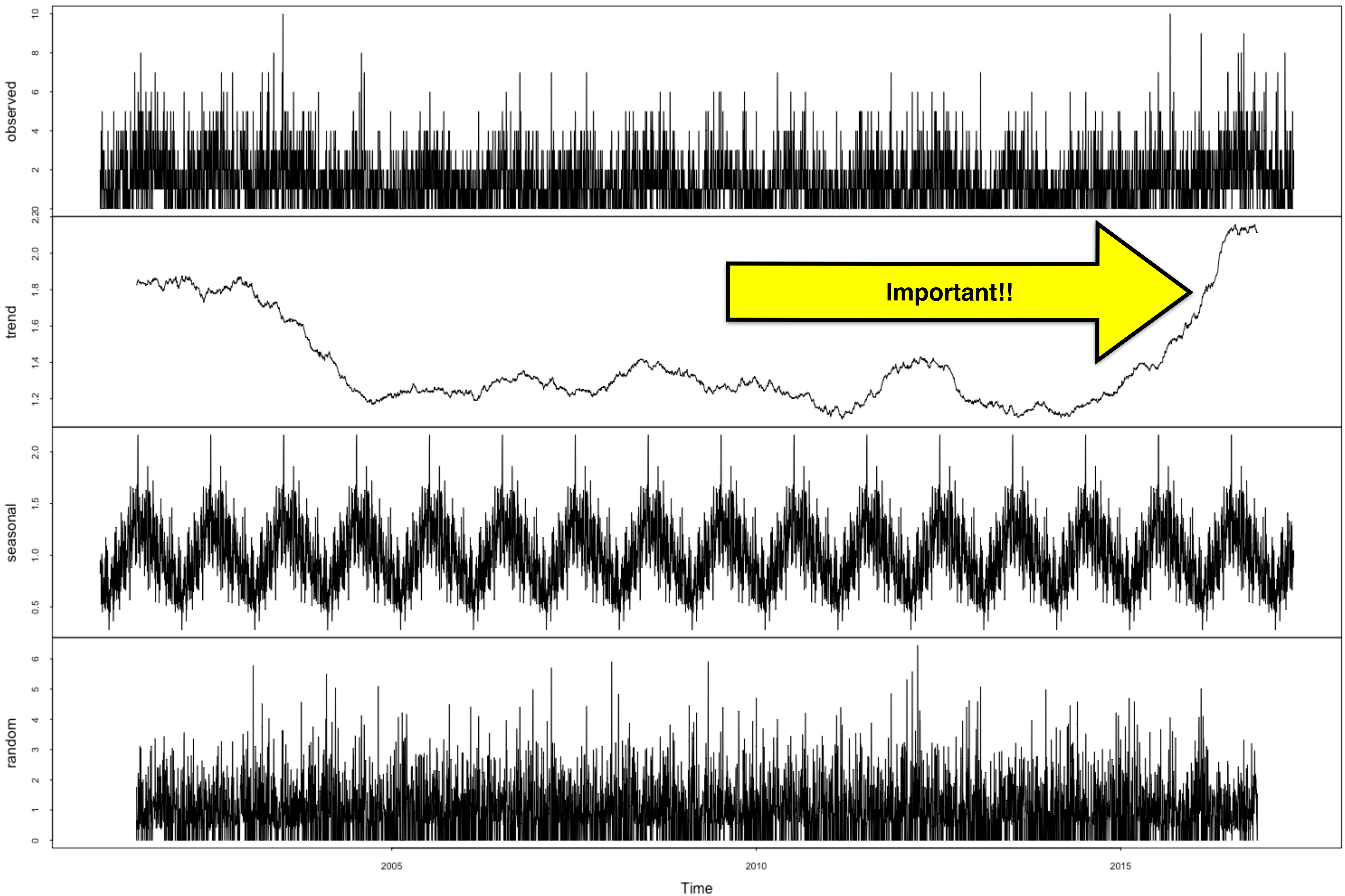








Time series decomposition of homicides in Chicago by day, January 1, 2001–May 14, 2017



Predicted Number of homicides in Chicago, by date and time series prediction method

	Naive	Meanf	RWF	SNaive	Unofficial Reported*
Sunday May 14, 2017	1	1.413	1	4	5
Monday May 15, 2017	1	1.413	1	4	0
Tuesday May 16, 2017	1	1.413	1	4	1
Wednesday May 17, 2017	1	1.413	1	1	4
Thursday May 18, 2017	1	1.413	1	2	2
Friday May 19, 2017	1	1.413	1	1	0
Saturday May 20, 2017	1	1.413	1	2	1
Sunday May 21, 2017	1	1.413	1	3	2
Monday May 22, 2017	1	1.413	1	1	3
Tuesday May 23, 2017	1	1.413	1	0	2
Wednesday May 24, 2017	1	1.413	1	1	1
Total	11	15.543	11	23	21
Accuracy vs unofficial reported	0.5238	0.7401	0.5238	1.0952	1

*https://www.dnainfo.com/chicago/2017-chicago-murders

Time Series Predictions for the other four crimes

Thefts	Naive	Meanf	RW	SNaive
Sunday May 14, 2017	155	220.524	155	161
Monday May 15, 2017	155	220.524	155	158
Tuesday May 16, 2017	155	220.524	155	166
Wednesday May 17, 2017	155	220.524	155	144
Thursday May 18, 2017	155	220.524	155	176
Friday May 19, 2017	155	220.524	155	158
Saturday May 20, 2017	155	220.524	155	169
Sunday May 21, 2017	155	220.524	155	166
Monday May 22, 2017	155	220.524	155	171
Tuesday May 23, 2017	155	220.524	155	175

Criminal Damage	Naive	Meanf	RWF	SNaive
Sunday May 14, 2017	89	121.8337	89	90
Monday May 15, 2017	89	121.8337	89	96
Tuesday May 16, 2017	89	121.8337	89	85
Wednesday May 17, 2017	89	121.8337	89	71
Thursday May 18, 2017	89	121.8337	89	105
Friday May 19, 2017	89	121.8337	89	94
Saturday May 20, 2017	89	121.8337	89	116
Sunday May 21, 2017	89	121.8337	89	117
Monday May 22, 2017	89	121.8337	89	81
Tuesday May 23, 2017	89	121.8337	89	74

Battery	Naive	Meanf	RWF	SNaiv
Sunday May 14, 2017	135	193.4009	135	134
Monday May 15, 2017	135	193.4009	135	175
Tuesday May 16, 2017	135	193.4009	135	141
Wednesday May 17,	135	193.4009	135	118
Thursday May 18,	135	193.4009	135	152
Friday May 19, 2017	135	193.4009	135	136
Saturday May 20,	135	193.4009	135	142
Sunday May 21, 2017	135	193.4009	135	169
Monday May 22, 2017	135	193.4009	135	182
Tuesday May 23, 2017	135	193.4009	135	131

Narcotics	Naive	Meanf	RWF	SNaive
Sunday May 14, 2017	30	115.6865	30	48
Monday May 15, 2017	30	115.6865	30	39
Tuesday May 16, 2017	30	115.6865	30	52
Wednesday May 17, 2017	30	115.6865	30	45
Thursday May 18, 2017	30	115.6865	30	46
Friday May 19, 2017	30	115.6865	30	82
Saturday May 20, 2017	30	115.6865	30	55
Sunday May 21, 2017	30	115.6865	30	37
Monday May 22, 2017	30	115.6865	30	48
Tuesday May 23, 2017	30	115.6865	30	44

Machine Learning Overview

Can we predict which high impact crimes based on the given inputs?

Assuming certain crimes incur greater costs to society we decided to isolate five high impact crimes in our study

- Homicide
- theft
- battery
- criminal damage
- narcotics
- all others

Machine Learning Models

- Random Forest
- Gradient Boosted
- Support Vector Machine
- Extreme gradient boosted
- Logistic Regression

Computational Limitation

Reference

Predicted

	Other	Battery	Criminal Damage	Homicide	Narcoticss	Theft	**PRECISION**
Other	**18,025**	8,027	5,377	131	1,757	8,579	**43.0%**
Battery	1,009	**1,049**	241	10	93	396	**37.5%**
Criminal Damage	374	213	**199**	3	23	166	**20.3%**
Homicide	9	5	1	**3**	4	3	**12.0%**
Narcotics	146	73	12	3	**223**	34	**45.4%**
Theft	931	367	307	0	32	**2,487**	**61.8%**
SENSITIVITY	**88.0%**	**10.8%**	**3.3%**	**2.0%**	**10.5%**	**21.3%**	

Accuracy: 43.8%	**Avg. Sensitivity: 22.6%**	**Avg. Precision: 36.7%**	

Machine Learning: Logistic Regression

The logistic regression training data was set up as 2001–2016 and the test data was set up as 2017. The data that could not be predicted (such as case number) was eliminated.

Confusion Matrices for the five logistic regression models:

Battery	0	1	% Acc
Down	69,354	0	1
Up	155	15,862	0.9903
Accuracy	0.991844		

Homicide	0	1	% Acc
Down	85,165	168	0.9980
Up	9	29	0.763
Accuracy	0.997926696		

Theft	0	1	% Acc
Down	65,806	0	1
Up	155	19,410	0.9921
Accuracy	0.9981844		

Criminal Damage	0	1	% Acc
Down	75,588	0	1
Up	155	9,628	0.9842
Accuracy	0.9981844		

Narcotics	0	1	% Acc
Down	81,959	3	0.9999
Up	9	3,400	0.9974
Accuracy	0.9998594		

Single computer full dataset

↓

Parallel processes with multicore

↓

Computer with more cores

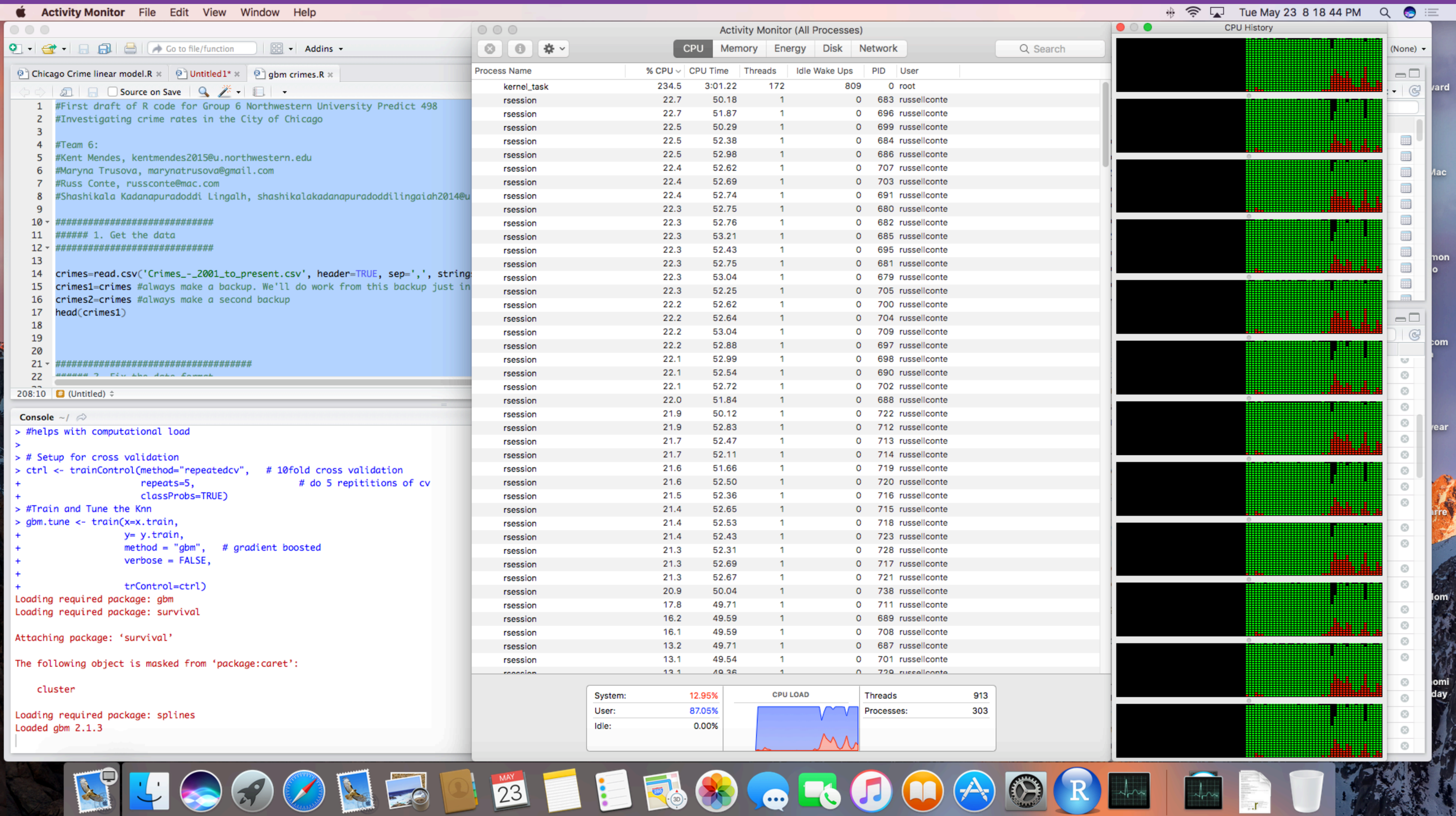
↓

Reduce Data Set

↓

Social Sciences Computer cluster

An example of a multi-core session in real time



ML Next Steps

- Extreme gradient on local computer
- Current jobs running (GBM, RF, ANN)
- Potential for increased accuracy with more inputs

A sample of what ML is seeing

Primary.Type	Arrest	Domestic	Beat	District	Ward	Community.Area	X.Coordinate	Y.Coordinate	Year	Latitude	Logitude
Other	TRUE	FALSE	612	6	17	71	1170652	1853593	2008	41.75374	–87.65019
THEFT	TRUE	FALSE	424	4	10	46	1197730	1845034	2008	41.72962	–87.55124
Other	FALSE	FALSE	2424	24	49	1	1162043	1948920	2008	42.01551	–87.67907
BATTERY	FALSE	FALSE	132	1	2	33	1177496	1891878	2008	41.85864	–87.62395
Other	FALSE	FALSE	1332	12	2	28	1162181	1900973	2008	41.88393	–87.67991
Other	FALSE	FALSE	524	5	34	53	1170471	1827663	2008	41.68259	–87.65180

- Add more variables to our data.
- Community areas would do well to study why some have less crime than the community area right next door to them
- Continue to refine the time series models to include locations
- Continue to deploy Tableau and other visualizations to help CPD and the Chicago community.
- Deploy police to educate the public on the data around crime and crime prevention
- Add support from the Mayor, City Council and all public officials to use data to reduce crime in Chicago