

Checkpoint 6: Visualization

MSAI 339

Prof. Jennie Rogers

December 2, 2018

GROUP 5

Jack R

Quincia H

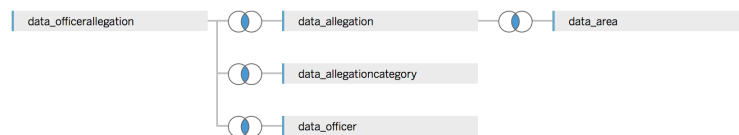
Ikhlas A

Purpose

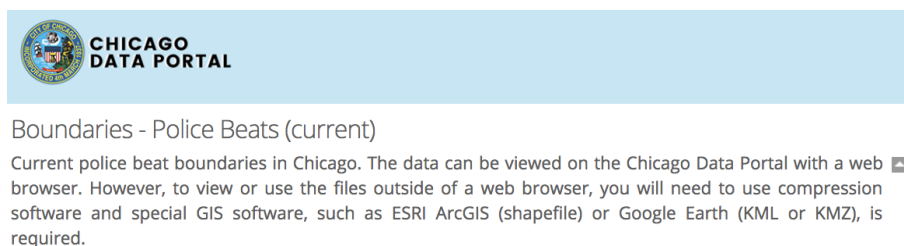
Being able to convey results and analysis in a way that is easy to understand is essential in data science. Data visualization allows for widespread understanding of result and knowledge. Tableau was leveraged for this checkpoint due to its straightforward user interface.

Tableau

Data Visualization sheets, dashboards, and stories are the bread and butter of Tableau. A unique feature of Tableau is the accessibility and data aggregation of multiple databases from different hosts. For this checkpoint, we connected Tableau to our Postgresql CPDB database as well as the crime data CSV file provided on the Chicago crime portal website. As in previous checkpoints, we will summarize our results and analysis of each question provided in the project proposal for visualization. To prepare the data to provide insight on the questions below, we used Tableau's built in data aggregation tool to build a tailored table that joined five of the cpdb tables into one.



Chicago Boundaries Dataset



To supplement our tables, we imported the *Police Boundaries Dataset* off of the *Chicago Data Portal*. Since the CPDB database provides beat ids and a table to grab the corresponding beat names, we used the boundaries dataset to connect each beat to their respective district. This let us increase the potential granularity of our distributions by area.

Visualization

It is important to note that the figures below do not encapsulate the interactive visualization that Tableau incorporates. As they may look cluttered in this format, the tooltip functionality

is a key aspect of the visualization.

Approach

An important factor of data visualization is to maximize the knowledge gain while minimizing the complexity of the graphic. Because a lot of our questions were related to location, our first thought was to chart the exact locations of crimes and complaints on a map chart. This resulted in a cluttered blob of data due to the sheer number complaints/crimes since 2001. Even with the extra dataset for boundaries and GIS polygons, plotting 200,000 allegation locations, let alone six million crime locations, proved to just clutter the chart. So, we focused on the different location attributes that were categorical.

What areas have the highest crime rate, and what areas exhibit the most complaints?

We decided to split this question up into numerous visualizations to be used in a dashboard. For both data sets, we evaluated crimes/allegations per district, beat, and rate over time. In addition, we provided the number of unique officers that received allegations in each district and the rate of change over time.

Crimes By District and Type

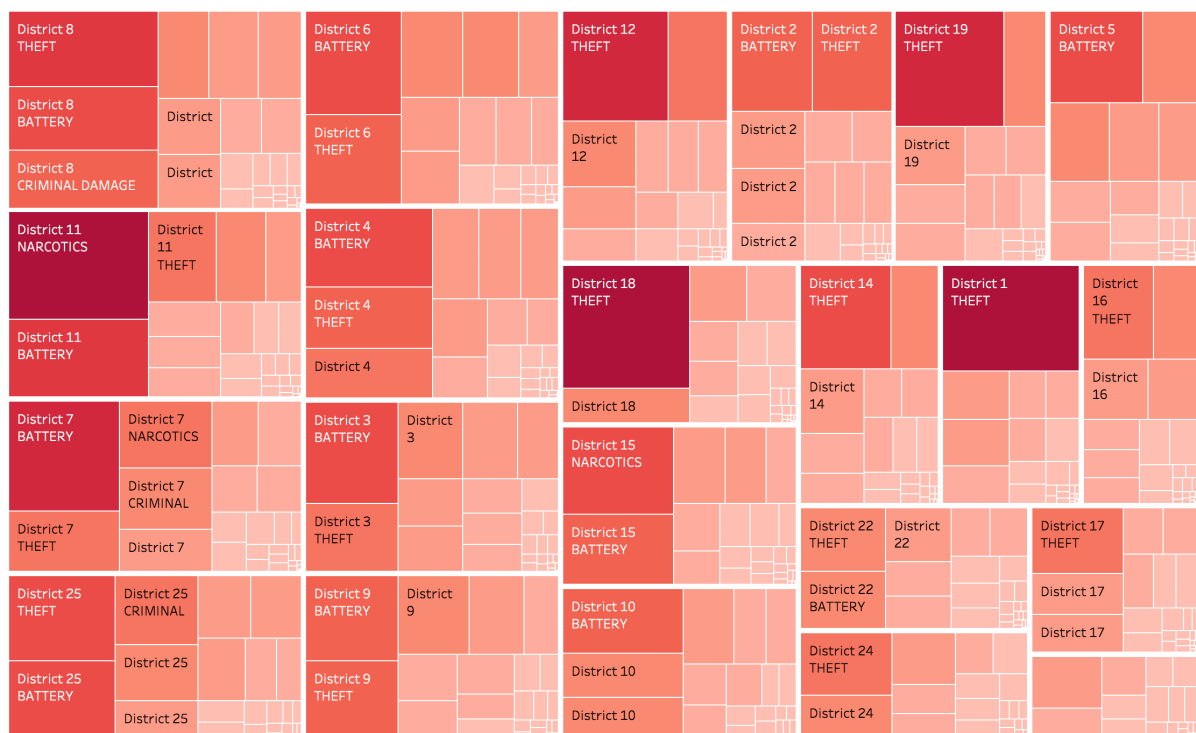


Figure 1: Highest Crime Rate and Most Common Crime

Allegations Per Beat/District

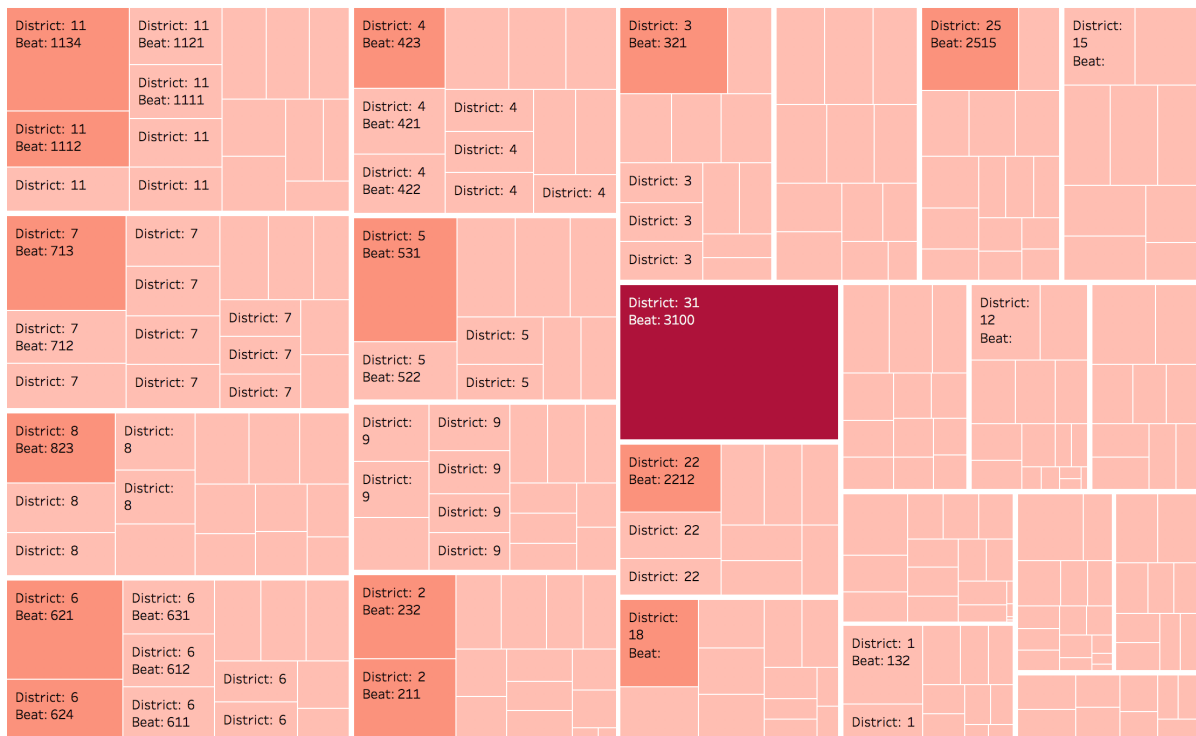


Figure 2: Allegations Per Beat Per District

How many unique officers have received complaints in each district or area of interest?

Allegations: Unique Officers Per District

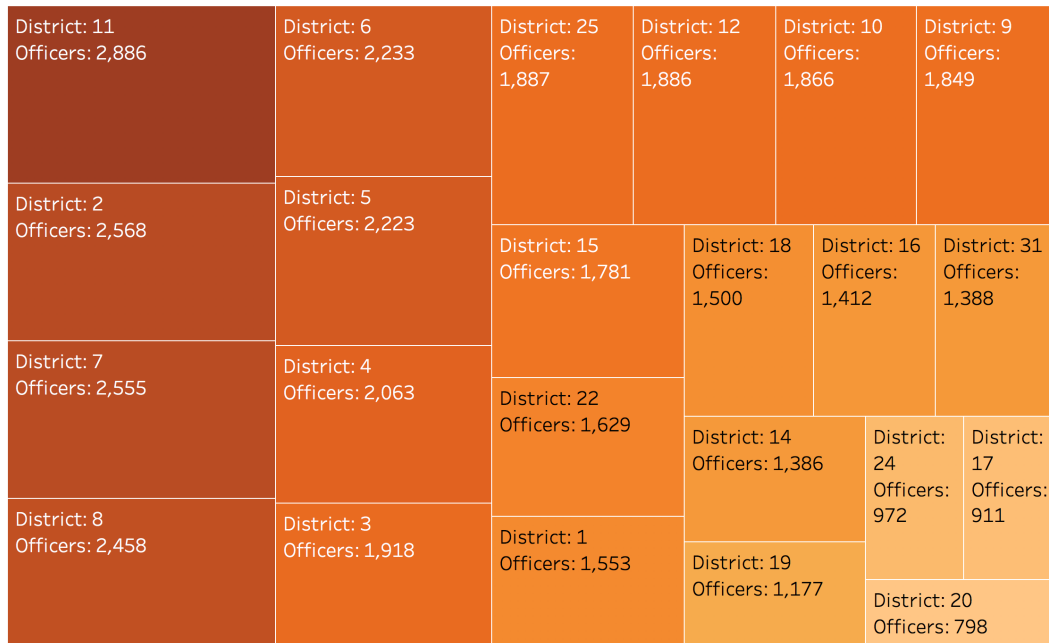


Figure 3: Unique allegation recipients per district

Allegations: Unique Officers Per Beat

Beat: 3100 1,388	Beat: 132	Beat: 725	Beat: 1533	Beat: 723	Beat: 1135	Beat: 1213	Beat: 1022	Beat: 324	Beat: 1033	Beat: 1123	Beat: 734	Beat: 411	Beat: 1414	Beat: 1623	Beat: 813	Beat: 934
	Beat: 624	Beat: 832	Beat: 434	Beat: 623												
Beat: 1134 1,163	Beat: 1112	Beat: 1531	Beat: 735	Beat: 2432	Beat: 424	Beat: 313	Beat: 715	Beat: 2533	Beat: 532	Beat: 731	Beat: 1012	Beat: 2234	Beat: 524	Beat: 1014	Beat: 632	Beat: 724
	Beat: 423	Beat: 1121	Beat: 825	Beat: 1115		Beat: 2535	Beat:									
Beat: 531 976	Beat: 712	Beat: 1524	Beat: 1024	Beat: 523	Beat: 312	Beat: 332										
Beat: 713 883	Beat: 1822	Beat: 512	Beat: 612	Beat: 835	Beat: 323	Beat: 431	Beat: 912									
Beat: 211 872	Beat: 421	Beat: 1111	Beat: 414	Beat: 2211	Beat: 932	Beat: 1131										
Beat: 621	Beat: 2212	Beat: 1122	Beat: 726	Beat: 213	Beat: 634	Beat: 2534	Beat: 714									
Beat: 232	Beat: 522	Beat: 1133	Beat: 733	Beat: 831	Beat: 331	Beat: 1511										
Beat: 2515	Beat: 511	Beat: 711	Beat: 513	Beat: 1021	Beat: 614	Beat: 2532	Beat: 824									
Beat: 321	Beat: 1224	Beat: 611	Beat: 1113	Beat: 925	Beat: 613	Beat: 1211										
Beat: 823	Beat: 1532	Beat: 631	Beat: 1522	Beat: 1011	Beat: 915	Beat: 1422	Beat: 433									

Figure 4: Unique allegation recipients per beat

How did the crime rate and allegation rate change over the years?

In the project proposal, our third question focused on visualizing showing the factors that went in to an allegation being filed. Due to how vague the question was, we changed it to tackle the derivative of the previous questions. By showing the rates of allegations and crimes on a district and beat level, it allowed us to build more concrete conclusions.

Number of Crimes Per District Per Year

District1	Date																
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Null				2	1	1	2	40									1
1	14,854	17,971	17,571	17,578	17,697	16,950	15,648	15,850	13,394	12,525	11,819	12,249	12,196	11,569	11,991	13,717	15,375
2	35,401	27,046	25,096	23,758	21,959	20,185	19,366	17,774	16,746	15,951	15,323	13,606	12,825	11,514	10,726	11,401	11,559
3	24,310	23,556	23,579	24,278	24,221	23,566	23,111	21,613	20,552	19,700	18,452	17,847	16,021	13,756	13,044	12,304	12,635
4	25,115	25,831	26,143	26,142	24,986	25,984	25,457	24,973	23,047	21,200	21,415	20,005	17,917	16,692	15,827	14,969	14,694
5	20,006	21,091	19,950	20,499	20,369	19,948	20,820	19,240	18,005	16,277	15,958	15,432	13,969	12,583	11,328	11,602	11,722
6	25,478	25,314	25,460	26,104	25,613	25,150	26,997	26,037	23,346	21,392	20,658	19,433	18,309	16,517	16,012	16,163	16,509
7	26,600	28,125	27,864	28,735	27,465	27,484	27,456	27,422	23,382	22,613	21,251	20,315	18,206	15,673	15,733	14,184	13,794
8	31,342	32,092	31,706	31,399	31,349	32,263	31,098	30,840	28,626	26,505	25,363	22,691	20,253	18,247	17,285	17,486	16,505
9	24,080	25,312	25,293	23,699	22,866	23,498	21,716	20,952	19,397	18,513	18,452	16,811	14,885	13,515	12,710	12,638	11,735
10	19,641	19,460	19,086	20,364	19,269	18,310	18,331	17,717	15,910	16,113	14,960	15,205	13,965	12,552	11,739	12,536	12,462
11	27,651	30,063	30,017	29,323	29,240	28,301	26,557	25,681	23,457	22,578	21,590	22,064	21,936	20,626	19,473	18,591	17,983
12	24,878	24,678	25,000	24,226	22,290	22,042	20,913	20,501	18,136	17,264	16,723	16,022	14,094	12,597	12,319	14,050	13,459
14	22,542	22,389	20,595	19,023	16,956	17,271	16,358	16,522	15,728	15,185	13,468	12,661	10,803	9,457	8,938	10,320	10,035
15	20,118	20,443	21,331	20,259	19,134	20,418	20,347	19,124	17,927	16,706	15,515	14,564	13,805	12,927	11,715	11,378	10,157
16	16,012	16,057	15,875	14,523	14,233	14,290	14,151	14,989	13,557	11,943	11,361	10,979	10,620	9,550	9,411	9,436	8,908
17	14,484	14,430	13,975	13,058	13,136	13,057	11,926	12,560	11,648	11,280	10,456	9,783	8,547	7,430	7,716	7,768	7,888
18	22,516	22,315	20,379	20,172	18,759	18,415	18,147	17,730	16,073	15,192	14,544	14,347	12,913	11,507	11,358	13,203	15,066
19	22,755	22,901	21,485	21,038	19,528	18,883	18,590	18,510	17,048	16,173	15,477	15,771	14,064	12,199	11,559	12,238	12,041
20	9,987	9,152	8,258	8,369	8,728	8,180	7,301	7,017	6,296	6,102	5,758	5,739	4,907	4,331	4,262	4,399	4,585
21			1	3													
22	14,952	15,674	14,959	15,438	15,110	14,861	14,980	14,292	13,549	12,764	11,686	10,903	10,276	8,801	8,732	8,585	8,323
24	15,110	15,020	13,933	14,396	14,323	14,186	12,993	12,848	11,846	11,256	10,209	9,618	8,737	7,549	7,023	7,341	8,177
25	27,918	27,825	28,370	26,990	26,468	24,849	24,719	24,785	24,975	23,036	21,291	19,877	17,784	15,616	15,031	14,501	14,209
31		8	11	6	5	7	15	21	20	25	22	4	5	3	9	14	2

Figure 5: Crime rate per district over time

Allegations Per District Over Time

District	Year of Incident Date																
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
1	255	224	213	143	128	130	162	146	127	136	157	154	163	149	123	28	10
2	742	532	425	411	234	303	349	355	330	330	330	236	279	181	198	62	17
3	376	443	393	342	278	305	381	438	437	304	366	345	343	296	231	56	8
4	570	691	539	486	300	506	628	506	505	414	379	380	298	356	191	103	16
5	526	548	421	410	373	352	532	511	420	341	341	356	331	344	198	94	13
6	595	628	513	437	294	338	593	522	575	574	438	480	399	286	270	70	27
7	666	734	765	734	461	539	648	606	608	595	526	460	406	338	273	56	26
8	622	687	737	543	311	517	532	486	581	516	412	319	308	297	215	65	19
9	478	542	496	386	261	365	434	441	510	413	307	267	213	219	137	30	15
10	374	459	413	371	308	388	363	384	336	328	262	277	211	215	211	33	22
11	723	892	701	674	478	680	796	683	607	597	428	411	398	456	265	117	19
12	333	366	282	206	170	193	153	216	168	138	123	136	167	160	122	44	17
14	347	381	307	263	189	199	329	247	235	173	167	166	111	74	60	10	12
15	343	438	342	387	326	367	480	377	341	314	269	196	274	180	136	49	10
16	235	247	230	177	199	206	219	239	202	166	166	125	131	101	123	33	19
17	162	168	162	143	92	166	198	170	148	115	92	82	66	64	69	13	8
18	397	494	367	289	271	227	324	248	211	211	140	167	155	104	128	47	8
19	202	158	183	144	89	107	169	168	119	119	126	151	127	143	110	38	9
20	144	159	121	112	125	81	115	103	84	96	84	61	68	40	70	17	1
22	320	450	401	323	205	270	364	266	369	271	236	186	166	201	151	70	5
24	327	297	227	168	212	190	273	239	199	160	139	93	98	91	72	17	8
25	418	407	357	313	290	284	398	406	434	322	260	225	218	235	168	58	22
31	315	342	321	345	354	285	354	354	516	234	204	183	219	114	132	51	3

Figure 6: Allegations per district over time

Results and Analysis

As stated above, visualization is very important as the purpose of data analysis is to convey a conclusion or a finding. Speculatively, our results produce concrete visualizations for each question. Although currently separated, the figures can be compared to identify positive and negative correlations. Showing total crimes by district would say district 11, 18, and 1 have the highest number of crimes. That being said, district 21 (incorrectly labeled 31 in CPDB), beat 3100 has the most allegations over the same time period. In figures 3 and 4, districts show a smaller range of values when compared to the distribution of unique officers by beat. This could mean that distributions by beat may provide more insight than by district because of the granular scale. One decision we struggled to handle was for question 3, and it corresponded to the lack of data in 2016-2018 for allegations. In figure 6, there is a large percentage decrease in allegations for 2016 and 2017 across all districts which could misinform viewers of an outrageous decrease in allegations and/or crimes due to incomplete data. Another minor change that could be made would be to remove the individual values off of figures 5 and 6.

Tableau - Pros and Cons

Tableau is a very powerful tool for a wide number of reasons ranging from data aggregation to the minimal learning curve that comes with interacting with it. With very little coding or database knowledge, you can quickly connect to a large variety of database hosts or even a local csv file and begin data exploration or visualization. The click and drag interactions are perfect for a first look at the data. The drawbacks of Tableau come from the limited number of charts you can encompass into your dashboards or visualizations. When compared to D3 or Vega, Tableau provides only a fraction of the charting capabilities.

Conclusion

This checkpoint enforced the importance of data visualization. It is common for data scientists to become engulfed by percentages and quantitative results without being able to understand the values or conclusions that could be drawn by visualizing them in a different way. Future work would be to expand off of the results we have acquired through Tableau by using D3 to build more complex abstractions.